



JRC TECHNICAL REPORTS

SEURAT-1 Tools & Methods Catalogue

*Report prepared by the European Commission's Joint Research Centre (JRC)
under the terms of contract No 32485-COLIPA-2011-T1CD ISP
between Cosmetics Europe and the JRC*

Version 1 (15 September 2016)

Editors:
Agnieszka Swiatek
Andrea Richarz
Annett Janusch-Roi
Elisabet Berggren

2016



EUR 28123 EN

This publication is a Technical report by the Joint Research Centre (JRC), the European Commission's science and knowledge service. It aims to provide evidence-based scientific support to the European policy-making process. The scientific output expressed does not imply a policy position of the European Commission. Neither the European Commission nor any person acting on behalf of the Commission is responsible for the use which might be made of this publication.

Contact information

Name: Elisabet Berggren

Address: Directorate F – Health, Consumers and Reference Materials

Chemicals Safety and Alternative Methods unit (F.3)

Via E. Fermi, 2749

I-21027 Ispra (VA)

Italy

E-mail: elisabet.berggren@ec.europa.eu

JRC Science Hub

<https://ec.europa.eu/jrc>

JRC102532

EUR 28123 EN

PDF ISBN 978-92-79-62118-5 ISSN 1831-9424 doi:10.2787/302272

Luxembourg: Publications Office of the European Union, 2016

© European Union, 2016

Reproduction is authorised provided the source is acknowledged.

How to cite: EC, SEURAT-1 Tools & Methods Catalogue; EUR 28123; doi:10.2787/302272

All images © European Union 2016 unless otherwise specified

Table of contents

Abstract.....	3
Introduction.....	5
SEURAT-1 Tools & Methods Catalogue	7
Research Initiative - SEURAT-1 - Towards the Replacement of <i>In Vivo</i> Repeated Dose Systemic Toxicity Testing (2011-2016).....	9
JRC Project.....	13
HepaRG repeated dose exposure assay of mitochondrial health and steatosis <i>Method Summary no 167.....</i>	<i>15</i>
ScrTox Project.....	23
Differentiation of induced-pluripotent stem cells into post-mitotic neurons and glial cells (mixed culture) <i>Method Summary n° 165.....</i>	<i>25</i>
Standard operating procedure for differentiation of human induced pluripotent stem cells into post-mitotic neurons and glial cells (mixed culture) <i>Protocol n° 165.....</i>	<i>33</i>
Standard operating procedure for expansion of rosette-derived neural stem cells <i>Protocol n° 166.....</i>	<i>45</i>
NOTOX Project.....	53
DNA methylation analysis using HumanMethylation450k Microarray <i>Method Summary n° 170.....</i>	<i>55</i>
DNA methylation analysis using HumanMethylation450 Microarray <i>Protocol n° 170.....</i>	<i>61</i>
Analysis of histone modifications via chromatin immunoprecipitation and sequencing (ChIP-seq) <i>Protocol n° 171.....</i>	<i>67</i>

COSMOS Project.....	75
Physiologically-Based Kinetic models	
<i>Method Summary n° 161.....</i>	<i>81</i>
The Virtual Cell Based Assay	
<i>Method Summary n° 162.....</i>	<i>103</i>
The Human Bio-Accumulation Model	
<i>Method Summary n° 163.....</i>	<i>119</i>
PPAR γ full agonists' pharmacophore-based predictor	
<i>Method Summary n° 168.....</i>	<i>131</i>
QSAR model for the prediction of LXR binding potential	
<i>Method Summary n° 169.....</i>	<i>137</i>
PPAR γ full agonists' potency predictor	
<i>Method Summary n° 175.....</i>	<i>145</i>
Parallel Artificial Membrane Permeability Assay (PAMPA) logPm predictor	
<i>Method Summary n° 176.....</i>	<i>151</i>
<i>In silico</i> screening for potential nuclear receptor ligands and alerts towards hepatosteatosis	
<i>Method Summary n° 177.....</i>	<i>157</i>
<i>In silico</i> profiler for covalent DNA binding	
<i>Method Summary n° 178.....</i>	<i>165</i>
<i>In silico</i> profiler for hepatotoxicity	
<i>Method Summary n° 179.....</i>	<i>177</i>
<i>In silico</i> profiler for mitochondrial toxicity	
<i>Method Summary n° 180.....</i>	<i>187</i>
<i>In silico</i> profiler for covalent protein binding	
<i>Method Summary n° 181.....</i>	<i>199</i>
<i>In silico</i> skin permeability estimation	
<i>Method Summary n° 182.....</i>	<i>211</i>

Abstract

The SEURAT-1 methods are here collected together and reported in the format of the database service that ensures OECD compliant descriptions. The information provided for each method is provided as reported and publicly available in the EURL ECVAM DataBase service on Alternative Methods (DB-ALM) developed and maintained by the Commission's Joint Research Centre (<http://ecvam-dbalm.jrc.ec.europa.eu>).

The aim of the FP7 SEURAT-1 Research Initiative is the development of a long-term research strategy leading to pathway-based human safety assessments in the field of repeated dose systemic toxicity testing of chemicals. The overall goal is to develop complementary theoretical, computational and experimental (in vitro) models that predict quantitative points of departure needed for human safety assessment and replace currently used animal testing.

Introduction

One of the most valuable outputs of SEURAT-1 is the portfolio of cutting-edge computational and *in vitro* tools and methods that will underpin new animal-free approaches to the safety assessment. Several of them were already being applied in the SEURAT-1 case studies to demonstrate their utility in predicting toxicity and supporting decision making.

The idea is that the *SEURAT-1 Tools & Methods Catalogue* only includes non-confidential information in a summarised and easy to read format – something that anyone can pick up and consult to see what is available. If there is interest in a particular method then more detailed information and generated data can be retrieved from the *ToxBank Data Warehouse* (<http://toxbank.net/>).

The SEURAT-1 Scientific Expert Panel agreed the catalogued SEURAT-1 methods should be disseminated via the public EURL ECVAM DataBase service on Alternative Methods, (DB-ALM) developed and maintained by the Commission's Joint Research Centre.

The DB-ALM is an online service (<http://ecvam-dbalm.jrc.ec.europa.eu/>) and provides standardised curated descriptions of over 250 alternative methods intended for use both in biomedical research and regulatory safety assessment and already includes information on methods developed and evaluated in other EU sponsored projects.

The SEURAT-1 methods are collected together and reported in the format of the database service that ensures OECD compliant descriptions. All methods included in the database, can be updated at any time, also after the projects are finalized. It is also assumed that the currently provided collection of methods is not the final outcome but that there still will be additional methods to be included originating from SEURAT-1. Some SEURAT-1 partners were rather reluctant to submit their methods to be included in the catalogue, partly due to the interest to publish methods and related outcome in peer-reviewed papers without pre-empting the results through the catalogue or any other publicly accessible database, but also because of the additional effort to provide summarised descriptions of the method.

In this report we have included the SEURAT-1 overview and the methods currently available in the Seurat-1 Catalogue.

SEURAT-1 Tools & Methods Catalogue

SEURAT-1 methods are intended for online dissemination through the DB-ALM as soon as new information is made available. An extract of the current status as of 15th September 2016 is included hereafter.

The compilation consists of method descriptions, contributed by each project. The dissemination medium are individual datasheets in pdf format, reviewed and compiled according to the harmonised content criteria, lay out and graphic style of the service.



Research Initiative - SEURAT-1 - Towards the Replacement of *In Vivo* Repeated Dose Systemic Toxicity Testing (2011 - 2016)

Systemic Toxicity

Study goal

The aim of the **SEURAT-1** Research Initiative is the development of a long-term research strategy leading to pathway-based human safety assessments in the field of repeated dose systemic toxicity testing of chemicals. The overall goal is to develop complementary theoretical, computational and experimental (*in vitro*) models that predict quantitative points of departure needed for human safety assessment and replace currently used animal testing.

Background

'Safety Evaluation Ultimately Replacing Animal Testing' (SEURAT) was presented as the target for chemical testing in the Seventh Framework Programme (FP7) under the Health team in 2008. The SEURAT Research Initiative started in January 2011 and is co-funded by the European Commission Directorate-General for Research & Innovation through FP7, and Cosmetics Europe with a total budget of EUR 50 million. The initiative was called **SEURAT-1**, indicating that this would be the first step in the specific area of repeated dose systemic toxicity addressing the global long-term strategic target of SEURAT to fundamentally change the way we assess the safety of chemicals, by superseding traditional animal experiments with a predictive toxicology that is based on a comprehensive understanding of how chemicals can cause adverse effects in humans.

Objectives

1. Formulate and implement a research strategy based on generating and applying knowledge of mode-of-action.
2. Develop highly innovative tools and methodology that can ultimately support regulatory safety assessment.
3. Demonstrate proof-of-concept at multiple levels, from theory to application.
4. Provide the blueprint for expanding the applicability domains - chemical, toxicological and regulatory.

Description

The **SEURAT-1** Research Initiative is designed as a coordinated cluster of five research projects, supported by a 'data handling and servicing' project (ToxBank) and a 'coordination and support' project (COACH). Detailed and updated information on the research initiative is available at <http://www.seurat-1.eu/>.



The projects of the cluster were:



Task: Stem cell differentiation for providing human-based organ specific target cells to assay toxicity pathways *in vitro*.

Coordinator: Marc Peschanski, INSERM / i-STEM, France

Website: www.scrtox.eu



Task: Development of a hepatic microfluidic bioreactor mimicking the complex structure and function of the human liver.

Coordinator: Catherine Verfaillie, KU LEUVEN, Belgium

Website: www.hemibio.eu



Task: Identification and investigation of human biomarkers in cellular models for repeated dose *in vitro* testing.

Coordinator: Jürgen Hescheler, Klinikum der Universität Köln, Germany

Website: www.detect-iv-e.eu



Task: Development of systems biological tools for organotypic human cell cultures suitable for long term toxicity testing and the identification and analysis of pathways of toxicological relevance.

Coordinator: Elmar Heinzle, Universität des Saarlandes, Germany

Website: www.notox-sb.eu



Task: Delivery of an integrated suite of computational tools to predict the effects of long-term exposure to chemicals in humans based on *in silico* calculations and estimation techniques.

Coordinator: Mark Cronin, Liverpool John Moores University, UK

Website: www.cosmos-tox.eu



Task: Supporting integrated data analysis and servicing of alternative testing methods in toxicology.

Coordinator: Barry Hardy, Douglas Connect, Switzerland

Website: <http://toxbank.net/>



Task: The coordination action, set up to optimise cooperation between the cluster projects and to pave the way for future research.

Coordinator: Bruno Cucinelli, ARTTIC, France

Each project had an individual Description of Work and a list of deliverables to fulfil. These deliverables are primarily contributing to the second SEURAT-1 objective on the development of new tools and methods.

SEURAT1 identified a set of standard reference compounds ('Gold Compounds'). They were chosen based on their rich database and known mode-of-action. They are divided in groups based on the target organ long-term toxicity. The three major groups being: hepatotoxic, cardiotoxic and renal toxic. More information on the compounds has been collected by ToxBank at the [Gold compounds wiki pages](#).

SEURAT-1 is also performing a series of proof-of-concept case studies on three different levels: theoretical (Adverse Outcome Pathways constructs), systems (developing integrated systems of testing methods to predict toxicity) and application (develop a quantitative safety assessment based on SEURAT-1 methods).

For more information on SEURAT-1, an annual report is issued each year of the research initiative (SEURAT-1, 2011-2016), which is downloadable at <http://www.seurat-1.eu/> or you can request a copy by writing to coach-office@eurtd.com.

Outcome

A valuable and easy accessible overview of the outcome of the SEURAT-1 projects will be collected in DB-ALM, searchable as the SEURAT-1 tools and methods catalogue.

Status

The SEURAT-1 started in 2011 and will be finalised at the end of 2016. The research activities were finalised at the end of 2015.

Remarks

Any updates of the SEURAT-1 tools and methods reported in DB-ALM will also be possible by SEURAT-1 partners after the end of the projects. Therefore, the documentation here collected is live and will not be finalised by the end of 2016, but will constitute a permanent source of continuously developed and refined alternative methods and related information for public access.

Sponsors

EUR 25 million funding are provided by the European Commission through FP7 HEALTH theme (DG for Research & Innovation) and EUR 25 million by Cosmetics Europe.

Contact

COACH office: ARTTIC, Sara Vinklatova (coach-office@eurtd.com)

Member of the COACH team at JRC: Elisabet Berggren, DG JRC (elisabet.berggren@ec.europa.eu).

Main References

[SEURAT-1 Annual Report, Eds. T. Gocht and M. Schwarz, Vol. 1, 2011.](#)

[SEURAT-1 Annual Report, Eds. T. Gocht and M. Schwarz, Vol. 2, 2012.](#)

[SEURAT-1 Annual Report, Eds. T. Gocht and M. Schwarz, Vol. 3, 2013.](#)

[SEURAT-1 Annual Report, Eds. T. Gocht and M. Schwarz, Vol. 4, 2014.](#)

[SEURAT-1 Annual Report, Eds. T. Gocht and M. Schwarz, Vol. 5, 2015.](#)

High resolution copies of the annual reports can be downloaded from the project's website:

<http://www.seurat-1.eu/pages/library/seurat-1-annual-report.php>

Last update: 19 May 2016

JRC

HepaRG repeated and single dose exposure study for mitochondrial health and lipid accumulation

This procedure was developed at JRC and evaluated in the Seurat-1 Research Initiative, launched by Cosmetics Europe and the European Commission with the long-term goal of achieving "Safety Evaluation Ultimately Replacing Animal Testing". It is a high throughput assay in 96-well plate format. It detects changes in mitochondrial membrane potential and lipid accumulation in of HepaRG cells, following repeated and single exposure to chemicals. High content imaging system was used for automated capture of fluorescence images and quantitative analysis. A detailed protocol is available as [JRC Technical Report Nr 98147](#).

Method Summary n° 167.....15

DB-ALM Method summary n° 167 : HepaRG repeated dose exposure in vitro assay of mitochondrial health and steatosis

Systemic Toxicity, Hepatotoxicity / Metabolism-mediated Toxicity

This high throughput experimental procedure detects changes in mitochondrial membrane potential and lipid accumulation in of HepaRG cells, following repeated and single exposure to chemicals. Fluorescently labeled live and fixed cells are analysed with high content imaging instrument. The procedure was developed as a part of the EU Integrated Project *Seurat-1*.

1. General Information

Corresponding author(s):

EURL ECVAM, European Commission - Joint Research Centre Italy

Date of data sheet creation: 27 April 2016

Last version number: 1

Last update: 27 April 2016

Assay Throughput:

High Throughput Assay in 96-well plate format. Platforms used: Microlab STARlet , Microlab Star (Hamilton, US), Cellomics ArrayScan® VTI (Thermo Scientific).

Status

Known Laboratory Use:

JRC (Italy)

Participation in Evaluation Studies:

This procedure was developed at JRC and evaluated in the Seurat-1 Research Initiative, launched by Cosmetics Europe and the European Commission with the long-term goal of achieving "Safety Evaluation Ultimately Replacing Animal Testing". More specifically, it is part of a SEURAT-1 Level 2 Proof of Concept case study, formalized with a research contract between Cosmetics Europe and the European Commission's Joint Research Centre (Contract No. 32485-COLIPA-2011-T1CD ISP).

2. Method Definition

Purpose of the Method

TYPE OF TESTING	: Screening, Part of an integrated testing strategy
LEVEL OF TOXICITY ASSESSMENT	: Hazard identification, Toxic potential
PURPOSE OF TESTING	: Screen or priority setting, Mechanistic studies

Context of use

This procedure was created for the SEURAT-1 Project to address liver toxicity endpoints relevant for the acute and repeated dose toxicity studies. Liver Adverse Outcome Pathways (AOPs) and their Key Events (reviewed in detail by Landesmann, 2015; and Vincken, 2015) have been used as guidance in selection of the most relevant endpoint assays. Selected assays were then adapted to the HTS platform to screen a set of reference chemicals across the selected endpoints. The ultimate goal of the study is to provide input data for high order analysis and modeling using cumulative data generated in the course of the Seurat-1 project.

Scientific Principle of the Method

Knowledge regarding molecular pathways leading to development of liver fibrosis, steatosis and cholestasis was used to carefully to select the endpoints indicative of potential liver toxicity in this assay.

These endpoints reflect either general organ (including liver) - toxic effects (such as cytotoxicity, mitochondrial damage, apoptosis) or liver-specific ones (such as accumulation of lipid droplets (steatosis)).

Human primary hepatocytes were used to study Cytochrome P450(CYP)-mediated metabolism and induction following exposure to new drugs in humans *in vitro*. However, their usage is limited by the restricted availability of liver tissue, their early phenotypic changes and their limited life-span. Furthermore, primary hepatocytes exhibit large variability at the level of CYP activities and the magnitude of induction after incubation with prototypical inducers - often greater than seen *in vivo* situation (Guillouzo et al., 1993). Therefore, for high throughput studies an established hepatic cell line would be more suitable.

The human hepatoma cell line HepaRG displays hepatocyte-like functions and expresses drug detoxifying enzymes at a relatively high levels drug transporter proteins and nuclear receptors, when compared to cell lines like HepG2 (Kanebratt et Andersson, 2008). HepaRG cells were obtained from a liver tumor of a female patient suffering from hepatocarcinoma and hepatitis C infection (Gripon et al., 2002). Undifferentiated bipotent proliferating HepaRG progenitor cells start to differentiate towards hepatocyte-like and biliary-like cells at confluence. If hepatocyte-like cells are selectively isolated and cultured at high cell density, they proliferate only slightly, while still preserving their differentiation status. However, when plated at low density, they transdifferentiate into hepatocyte and biliary lineages through a bipotent progenitor stage (Cerec et al., 2007)

High content cell imaging can be used to detect changes in mitochondrial membrane potential and lipid accumulation following the exposure of HepaRG cells in a wide range of treatments on a high throughput platform. The treatment scheme includes both single and repeated exposure scenarios to approximate the effects of sub-chronic chemical exposure.

The assay is designed to observe cells that are still viable, (and therefore attached to the surface of the well), but undergoing physiological changes due to chemical exposure. To achieve this, a combination of four fluorescent molecular probes with non-overlapping excitation and emission spectra are used.

Procedure Description

Biological Endpoint and Measurement:

The method assesses four biologically relevant endpoints:

MITOCHONDRIAL FUNCTION is evaluated with the Mitochondrial Health Kit (Invitrogen, H10295), measuring depolarization of mitochondrial membrane potential, which is considered a good indicator of mitochondrial dysfunction.

CELL VIABILITY is evaluated with the Image-iT DEAD Green viability stain, which is included in the Mitochondrial Health Kit. The stain is a marker of membrane permeability and its purpose is to differentiate between healthy cells and those whose physiology due to toxic exposure has already been slightly compromised.

LIPID ACCUMULATION is quantified with the LipidTOX Deep Red neutral lipid stain (Molecular probes, H34477) which has an extremely high affinity for neutral lipid droplets, considered indicators of chemically induced steatosis in HepaRG cells.

CELL NUMBER in each well is estimated by counterstaining fixed cells with Hoechst 33342 nucleic acid stain. The fluorescence is directly proportional to the number of nuclei in the sample.

Endpoint Value:

SIGNAL TO NOISE RATIO (S/N) is calculated for each chemical and each concentration tested

TRESHOLD value is used to assess the significance of Signal to Noise (S/N) ratio. E.g. threshold of -3 or less (for downwards effect) and +3 or more (for upwards effect) indicates a significant effect.

EC₅₀ - the half maximal effective concentration of a test substance

Experimental System:

Monolayer culture of human hepatoma cell line HepaRG™. The cells are a human hepatic progenitor cell line that retains many characteristics of primary human hepatocytes. They are patented (PCT/FR02/02391 of July 8, 2002) and can only be used under MTA from a licensed distributor.

Metabolic Competence:

Experimental system is metabolically competent

Exposure regime:

Application type : Repeated dose

Application volume : 85 µl

Dosage : Series of 11 dilutions by factor 1:3.
Chemical applied once or refreshed every 6 hours

Min : 0

Max : 250µM

Exposure time :

Min : 4 hours

Max : 72 hours

Controls used:*Positive Controls (PCs):*

Cell viability (Dead Green stain) - Valinomycin 7.2 µM

Mitochondrial toxicity (MitoHealth stain)– Valinomycin 165 nM

Lipid droplets formation (LipidTOX stain) – Cyclosporine A 30 µM

Negative Control (NC):

DMSO 0.1 % (added to four wells in every 96-well plate)

Basic procedure:

Main steps of the procedure are outlined in the **Figure 1** below. Briefly: 72 hours after the seeding, HepaRG cells are exposed to a set of 90 test chemicals and 2 positive controls at 11 concentrations (dilution 1:3). In each plate, 4 negative control wells (cells treated with only DMSO) are included.



Figure 1: Outline of the main steps of the procedure

The cells are incubated at 37°C, 5% CO₂, 100% humidity for 24 or 72 hours. The medium, containing the chemicals is refreshed every 6h. All the treatment procedures are fully automated using the STAR platform (Hamilton) and the entire process is managed by Hamilton software.

After the exposure cells are stained with Mitochondrial Health kit for 30 min. Afterwards the cells are fixed with 4% Formaldehyde and counterstained with Hoechst 33342 for 15 min. After fixation the cells are washed with PBS once and LipidTox staining solution is added for 30 min. The plates are stored overnight at +4°C, prior to high content imaging with Cellomic ArrayScan® (Thermo Scientific). The image acquisition is done using a set of four filters, selective for each fluorophore. Further details are made available in the JRC Technical Report No. 98147 (Jossens et al., 2015).

Quality/Acceptance Criteria**Culture conditions**

All procedures must be carried out under aseptic conditions and in the sterile environment of a laminar flow cabinet (class II biological hazard standard). All methods and procedures must be adequately documented and raw data archived for historical reference.

Undifferentiated HepaRG cells are used as starting culture in this assay. The cells must have to be free of contamination and cannot be passaged more than 20 times. At passage 21 the cells will go to senescence.

Chemical solubility

Poor chemical solubility can significantly affect the results of a toxicological study. To identify potential solubility issues in an objective and quantitative manner, the JRC procedure uses a nephelometric analysis (see page 15 of the JRC Technical Report). Highest concentration of a compound in DMSO is 250 mM. If there is precipitation, serial dilutions are prepared until no precipitation is observed. These solutions, and further 1:1000 dilutions in the culture medium are then tested by nephelometer. Solutions with turbidity level higher than 10 nephelometric turbidity units are regarded as insoluble, and necessitate further dilution of both stock and working solutions. For each chemical, the highest soluble concentration (used for the testing) is recorded in Annex 2 of the JRC Technical Report No. 98147 (Jossens et al., 2015).

Control plates

To ensure the quality of each run several controls are included: solvent control and the positive controls for the endpoints measured (cell viability, mitochondrial potential and lipid accumulation). For a run to be considered valid each normalised positive control should produce a clear dose-response curve, the replicate curves should overlap and there is a statistically significant effect at the relevant concentrations. In particular the cells treated with the Valinomycin dilution range starting from 165 μ M should display a significant decrease of mitochondrial potential but no detectable loss in viability or cell number.

HTS specific issues

In some runs, certain plates showed a position effect i.e. the behaviour of cells depends on where the well was located on the plate. In order to remove this effect, an additional mathematical approach can be applied and it should be documented in the data analysis files. Another approach used at JRC to counter the positional effects was to randomly rotate the position of the negative control.

Data Interpretation and/or Prediction Models

Data Analysis:

Normalisation

Raw data of multiple read outs are captured for all 96-well plates containing cells treated with chemicals, solvent controls and positive controls. Further processing is done in Excel as follows: an average of the negative control per each plate (\overline{NC}_c) and standard deviation of all negative controls across all plates (SD_{nc}) are calculated and used to normalise the read out for each well.

Interpretation

For each chemical the normalised data are plotted against the concentration tested and the confidence interval and the strength of the observed effect is calculated. Note that for the endpoints such as mitochondrial depolarisation the observed effects for the chemicals tested at the JRC were both positive and negative.

This part of the assay is still in development and current status is presented in detail in the JRC Technical Report No. 98147 (Jossens et al., 2015).

Prediction Model (PM):

Not available.

3. Method Performance

Robustness of the Method

Within-laboratory reproducibility:

4 runs with 90 chemicals each were conducted at the JRC. The method was judged by the developers as highly reproducible. The results are now being analysed in the preparation of the final Seurat-1 case study (Joossens et al., 2015).

Test Compounds

90 chemicals with well documented mode of action and present or absent hepatotoxic effects were tested in the JRC study and were mostly pharmaceuticals (selection was based on the work of O'Brien et al., 2006; Xu et al., 2008; Gomez-Lechon et al., 2010; Porceddu et al., 2012 and Garside et al., 2014). The complete list, including the highest soluble concentrations is included in the Annex 2 of the JRC Technical Report No. 98147 (Jossens et al., 2015).

Applicability:

The method is suitable for testing of pure solid and liquid organic and inorganic chemicals provided that they are soluble in both DMSO and culture medium to a sufficient extent. It is not suitable for volatile substances.

4. Discussion

Ethical issues and considerations for 3R's impact

The ultimate goal of the Seurat-1 Research Initiative is to develop complementary theoretical, computational and experimental (*in vitro*) models that predict quantitative points of departure needed for human safety assessment and replace currently used animal testing in the field of repeated dose systemic toxicity testing of chemicals.

The hepatotoxicity was one the main endpoints chosen by the consortium to study in depths as it is known to be a major risk in the drug development and prolonged chemical exposure. The predictive capacity of the animal models is often limited and the AOP-based testing strategies gain momentum as evidence-based and sound scientific approaches to the problem (Vinken, 2015).

The "*HepaRG repeated dose exposure assay of mitochondrial health and steatosis*" uses a well established human cell line HepaRG which, when differentiated, displays hepatocyte-like characteristics including morphology, expression of key metabolic enzymes, nuclear receptors and drug transporters (see e.g. <http://www.heparg.com>).

Potential for further development

The "*HepaRG repeated dose exposure assay of mitochondrial health and steatosis*" was developed with the aim of providing information that can be used for detection and classification of potential liver toxicity. It was a first a study of this kind, combining the currently best available *in vitro* model of liver cells, the high throughput approach and physiologically relevant endpoints. Preliminary analysis of the data revealed that whereas the methods is very reproducible and sensitive with the positive controls. However matching the hepatotoxic effects of reference chemicals with the *in vitro* test results presented a challenge. More data are needed from more chemicals or by including information of other relevant events of the AOP by either expanding the assay with probes specific for other such as oxidative stress, apoptosis, vesicle transport, or pooling the test results with datasets obtained with another method such as CYP induction or transcriptomic studies.

Other methods for assessing hepatotoxicity

In 2008 DB-ALM has compiled a thematic review of "*Hepatotoxicity/ metabolism-mediated toxicity*", intended as an overall view about the information content on alternative methods provided at the time the DB-ALM for the entire topic area (DB-ALM, 2008).

Another recommended extensive overview of the *in vitro* systems used to predict human hepatotoxicity has been published by Hengstler et al. (2014). Here the authors discuss main available hepatocyte *in vitro* systems and their possibilities as well as limitations in studying hepatotoxicity and ADME.

5. Contact Details

EURL ECVAM

Chemicals Safety and Alternative Methods Unit - European Union Reference Laboratory for Alternatives to Animal Testing (EURL-ECVAM)

European Commission - Joint Research Centre
 Corresponding author
 via E. Fermi 1
 Ispra (VA) 21027
 Italy
 email: ecvam-contact@jrc.ec.europa.eu
 telephone: 39 0332 7813527

Elisabeth Joossens

Senior statistician
 Chemicals Safety and Alternative Methods Unit - European Union Reference Laboratory for Alternatives to Animal Testing (EURL-ECVAM)
 European Commission - Joint Research Centre
 Method developer / user
 Via Enrico Fermi 2749
 Ispra 21027
 Italy
 email: Elisabeth.JOOSSENS@ec.europa.eu
 telephone: +39 0332 786785

Dr Tomislav Horvat

Research fellow
 Chemicals Safety and Alternative Methods Unit - European Union Reference Laboratory for Alternatives to Animal Testing (EURL-ECVAM)
 European Commission - Joint Research Centre
 Method developer / user
 via E. Fermi 1
 Ispra (VA) 21027
 telephone: 39 0332 7813527

Ver n°	Date	Name	Type of change
1	27 April 2016	EURL ECVAM	Datasheet compiled on the basis of the JRC Report No. 98147

6. Abbreviations & Definitions

AOP: Adverse Outcome Pathway

HTS: High Throughput Screening

JRC: Joint Research Center

NC: Negative control

PC: Positive control

SEURAT-1: FP7 EU Integrated Project "*Safety Evaluation Ultimately Replacing Animal Testing*". The aim of the SEURAT-1 Research Initiative is the development of a long-term research strategy leading to pathway-based human safety assessments in the field of repeated dose systemic toxicity testing of chemicals. Project Website: <http://www.seurat-1.eu/>

7. Bibliography

Method development:

- Joossens E., Bories G., Cole T., Fortaner Torrent S., Gineste J-M., Horvat T., Canovas Jorda D., Lostia A., Macko P., Novak J., Palosaari T. and Whelan M., (2015)
 JRC Technical Report No. 98147: Seurat-1: HepaRG, repeated and single dose exposure for Mitochondrial Health and LipidTox
[JRC publication repository](#) EUR 27756 EN

General references:

- Cerec, V., Glaise, D., Garnier, D., Morosan, S., Turlin, B., Drenou, B., Gripon, P., Kremsdorf, D., Guguen-Guillouzo, C., and Corlu, A. (2007)
Transdifferentiation of hepatocyte-like cells from the human hepatoma HepaRG cell line through bipotent progenitor.
Hepatology 45, 957–967
- DB-ALM (2008)
Hepatotoxicity / Metabolism-mediated Toxicity - Review Document
[*EURL ECVAM Database service on Alternative Methods to animal experimentation*](#)
- Garside, H., Marcoe, K.F., Chesnut-Speelman, J., Foster, A.J., Muthas, D., Kenna, J.G., Warrior, U., Bowes, J., and Baumgartner, J. (2014)
Evaluation of the use of imaging parameters for the detection of compound-induced hepatotoxicity in 384-well cultures of HepG2 cells and cryopreserved primary human hepatocytes.
Toxicology In Vitro 28, 171–181
- Gripon, P., Rumin, S., Urban, S., Le Seyec, J., Glaise, D., Cannie, I., Guyomard, C., Lucas, J., Trepo, C., and Guguen-Guillouzo, C (2002)
Infection of a human hepatoma cell line by hepatitis B virus.
Proceedings of the National Academy of Sciences of the United States of America 99, 15655–15660
- Guillouzo, A., Morel, F., Fardel, O., and Meunier, B. (1993)
Use of human hepatocyte cultures for drug metabolism studies.
Toxicology 82, 209–219
- Gustafsson, F., Foster, A.J., Sarda, S., Bridgland-Taylor, M.H., and Kenna, J.G. (2014)
A correlation between the in vitro drug toxicity of drugs to cell lines that express human P450s and their propensity to cause liver injury in humans.
Toxicological Sciences 137, 189–211
- Gómez-Lechón M.J., Tolosa L., Castell J. V. & Teresa M. (2010)
Mechanism-based selection of compounds for the development of innovative in vitro approaches to hepatotoxicity studies in the LIINTOP project.
Toxicology In Vitro 24, 1879–1889
- Hengstler, J., Hammad, S., Ghallab, A., Reif, R., and Godoy, P. (2014). , pp. . (2014)
In Vitro Systems for Hepatotoxicity Testing
In Vitro Toxicology Systems, 27–44
- Kanebratt, K.P., and Andersson, T.B. (2008)
HepaRG cells as an in vitro model for evaluation of cytochrome P450 induction in humans.
Drug Metabolism and Disposition 36, 137–145
- Landesmann, B. (2015)
AOP:38: Protein Alkylation leading to Liver Fibrosis <https://aopwiki.org/wiki/index.php/Aop:38> (Last Access: 26.04.2016)
- O'Brien, P.J., Irwin, W., Diaz, D., Howard-Cofield, E., Krejsa, C.M., Slaughter, M.R., Gao, B., Kaludercic, N., Angeline, A., Bernardi, P., et al. (2006)
High concordance of drug-induced human hepatotoxicity with in vitro cytotoxicity measured in a novel cell-based model using high content screening.
Archives of Toxicology 80, 580–604
- Porceddu, M., Buron, N., Roussel, C., Labbe, G., Fromenty, B., and Borgne-Sanchez, A. (2012)
Prediction of liver injury induced by chemicals in human with a multiparametric assay on isolated mouse liver mitochondria.
Toxicological Sciences 129, 332–345
- Vinken, M. (2015)
Adverse Outcome Pathways and Drug-Induced Liver Injury Testing.
Chemical Research in Toxicology 28, 1391–1397
- Xu, J.J., Henstock, P.V., Dunn, M.C., Smith, A.R., Chabot, J.R., and de Graaf, D. (2008)
Cellular imaging predictions of clinical drug-induced liver injury.
Toxicological Sciences 105, 97–105

Last update: 27 April 2016



Differentiation of induced-pluripotent stem cells into post-mitotic neurons and glial cells (mixed culture)

The neural differentiation protocol for human induced pluripotent stem cells (hiPSCs) has been developed at the JRC. In the course of the [Scr&Tox project](#) it has been optimised for application to mechanistic studies of the toxicity pathways triggered by the exposure to a chemical.

Method Summary n° 165.....	25
Protocol n° 165.....	33
Protocol n° 166.....	45

DB-ALM Method summary n° 165 : Differentiation of induced-pluripotent stem cells into post-mitotic neurons and glial cells (mixed culture)

Stem cell culture method, Systemic Toxicity, Neurotoxicity

This method describes all the stages of differentiation of human induced-pluripotent stem cells into neural precursor cells and further into mixed cultures of post-mitotic neurons and glial cells.

1. General Information

Corresponding author(s):

Dr Anna Price, European Commission - Joint Research Centre

Date of data sheet creation: 05 June 2015

Last version number: 1

Last update: 05 June 2015

Proprietary and/or Confidentiality Issues:

For this method the IMR90 cell line (fetal lung fibroblasts) was reprogrammed towards pluripotency generating human induced pluripotent stem cells (hiPSCs) by I-Stem (Evry, France) and kindly provided. On 8th October 2012 a material transfer agreement was signed between I-Stem and JRC where the hiPSC (IMR90) cell culture has been made available to JRC within the frame of Scr&Tox project, for non-profit research purposes.

Assay Throughput:

Not applicable

Status

Known Laboratory Use:

EURL ECVAM

Participation in Evaluation Studies:

The neural differentiation protocol for human induced pluripotent stem cells (hiPSCs) has been developed at the JRC and applied to mechanistic studies of the toxicity pathways triggered by the exposure to a chemical.

Pistollato et al. (2014) has developed a novel hiPSC-based *in vitro* model and assay for measuring the perturbations of the CREB pathway which is crucial in the development of the central nervous system (CNS), including neuronal survival, neurite outgrowth, precursor proliferation and neuronal differentiation (Lonze et al., 2002).

The hiPSC-derived neuronal cultures are considered as useful tools to study chemical-induced effects on developing human neuronal cells, providing a valuable alternative to the use of animal *in vitro* cell models, including cancer cell lines that often do not reproduce physiological cell biology (Pistollato et al., 2014).



In the course of the Scr&Tox project (<http://www.scrtox.eu/>) two SOP's were developed for the expansion and differentiation of the hiPSC-derived neural stem cells (NSCs), available from DB-ALM as:

- **DB-ALM Protocol n° 165: Standard operating procedure for differentiation of human induced pluripotent stem cells into post-mitotic neurons and glial cells (mixed culture)**
- **DB-ALM Protocol n° 166: Standard operating procedure for expansion of rosette-derived neural stem cells**

2. Method Definition

Purpose of the Method

TYPE OF TESTING : Replacement (partial), Part of an integrated testing strategy

Context of use

Currently the method is used for **Non-regulatory** purposes, mainly for basic research.

This method summary describes the procedure for differentiation of the hiPSCs into neural precursors (NSCs) and further into mixed cultures of post-mitotic neuronal and glial cells.

Human pluripotent stem cells (PSCs) offer a unique opportunity to develop a wide variety of human cell-based test systems because they may be expanded indefinitely and triggered to differentiate into any cell type. The differentiation of hiPSCs into mixed neuronal/glial culture and the establishment of stem cell-based assays for neurotoxicology could provide proof of concept on the use of these cell models for identifying relevant "toxicity pathways" which result in adverse health effects.

Neither the exposure regime nor endpoint measurements are currently defined in both DB-ALM Protocols n° 165 and 166 and are based on case by case studies (fit for purpose).

Scientific Principle of the Method

hiPSCs are widely used in the areas of basic research, regenerative medicine and toxicology. They are considered as a powerful tool for drug screening based on *in vitro* testing strategies. This expertise can be further applied for the development of new *human in vitro* tests which address current challenges in safety assessments. In particular, those related to different susceptibility to toxicants in human populations as stem cells can be obtained for individuals, leading to personalised medicine and toxicity studies. It allows avoiding interspecies variations since extrapolation of animal data is not required. Additionally, hiPSCs might in the future replace human embryonic stem cell-derived cultures, as they do not represent an ethical and/or legislative challenge for application in modern toxicity assays (Pistollato et al., 2012; Canovas-Jorda et al., 2014; Grskovic et al., 2014).

A major challenge for the application of stem cell-based models in toxicology and drug discovery is to demonstrate that they can reproducibly generate a significant range of toxicologically relevant cell types and that these cells provide characteristics and responses typical of human tissue types (Trosko and Chang, 2010). The general problem is that cells derived from iPSCs often do not differentiate and, for this reason, further effort is required to optimise existing differentiation protocols in order to be able to obtain functionally and terminally differentiated cells (including the neuronal lineage).

Amongst the differentiated cell types that can be obtained from PSCs and used for toxicology studies, neurons have been the most extensively studied (Kumar et al., 2012). HiPSC cells can be indefinitely expanded and efficiently differentiated into neuronal derivatives, including different regionalized neuronal subtypes, glial cells and peripheral neurons.

Neuronal models derived from hiPSCs are particularly suitable for developmental neurotoxicity testing, including application within a battery of tests for a chemical-induced neural teratogenicity. However before any testing is performed, these cultures have to be carefully characterized at cellular, molecular and functional level in order to define how they reflect the functionality of the *in vivo* central and peripheral nervous systems (Trosko and Chang, 2010; Pistollato et al., 2012).

The differentiated cell population should be shown to express at an appropriate level specific neural related markers, such as β -III-tubulin, microtubule-associated protein 2 (MAP2), neurofilament 200 (NF200) and Synapsin-I for mature neurons (Pistollato et al., 2012 and 2014). Once a marker panel has been selected, specific quality control methods are needed to establish acceptability criteria, which should also include cell functionality. Moreover, PSC-derived neurons should be proven to be electro-physiologically active and thus, specific quality control metrics for the functional activity and threshold levels for a positive phenotype need to be defined in order to properly judge the suitability of these cells for use in a toxicology assay.

The demonstration of the reliability and relevance of PSC-based *in vitro* toxicity tests for a given purpose is mandatory for the use of PSCs in regulatory toxicity testing. A high level of standardisation of the neuronal differentiation process is required in order to ensure the establishment of robust test systems

(Pistollato et al., 2012).

Procedure Description

Experimental System:

hiPSs (IMR90-hiPSC culture) - Human induced-pluripotent stem cells (hiPSC) derived from human foetal lung fibroblast (IMR90), reprogrammed towards pluripotency by retroviral transduction of Oct4 and Sox2, using the pMIG vectors. Created and distributed by I-Stem (Evry, France).

Metabolic Competence:

Not known

Exposure regime:

Application volume : Adjusted for each case study
 Dosage : Adjusted for each case study
 Exposure time : Adjusted for each case study

Basic procedure:

The current method describes the differentiation of IMR90-hiPSCs into post-mitotic neurons and glia. The cells are seeded at appropriate density and medium and cultured for 10 days. The cells begin to form colonies which are cut out and plated on low attachment petri dishes where they form embryoid bodies (EBs) floating in the medium. In a further stage, EBs are plated on poly-ornithine/laminin-coated plates and in the presence of the neural induction medium, and begin to generate visible rosettes after 5-8 days. Finally, rosettes are cut out, resuspended in neural differentiation medium and cultured for 21 days, where they generate post-mitotic neuronal and glial cells (see **Figure 1** below).

Figure 1: Representative scheme summarizing the described neuronal differentiation protocol.

Quality/Acceptance Criteria

Good Cell Culture Practice (GCCP) principles are recommended (Coecke et al., 2005; Pistollato et al., 2012).

3. Method Performance

Robustness of the Method

General performance measures:

The current method summary accompanies a robust and relatively rapid protocol: **DB-ALM n°165 : Standard operating procedure for differentiation of human induced pluripotent stem cells into post-mitotic neurons and glial cells (mixed culture).**

The protocol is developed for the neural differentiation of hiPSCs into heterogeneous cultures of mixed post-mitotic neurons (GABA-ergic, glutamatergic and dopaminergic neurons) and also glial cells (astrocytes, oligodendrocytes and microglia). The latter are fundamental for myelin formation, for the physiological functionality of neuronal networks and for neuronal homeostasis (Ghosh et al., 2011).

The reproducible differentiation protocol, yielding heterogeneous cultures of relevant cell types, including progenitors and terminally differentiated cells, is necessary to predict physiological *in vivo* cellular responses to toxicant exposure (Trosko and Chang, 2010). Indeed, this cell model could be used to perform a wider screening of chemicals, toxic not only for neurons.

The obtained neuronal cultures are suitable for studying different neuronal-related signalling pathways, such as the cyclic AMP response element binding (CREB) protein pathway (Pistollato et al., 2014). When strong perturbations of these pathways occur as a consequence of toxicant exposures, these pathways

can be considered as "toxicity" pathways. Pathway perturbations can be monitored by mean of cell-specific (neuronal and glial) physiological and functional readouts (e.g., immunocytochemical analysis and high content imaging, qPCR, multi-electrode array (MEA) analysis).

Moreover, there could be an association of the neuronal cultures with chemical perturbations of signalling pathways that are studied from other research groups in order to evaluate the relevance of selected "toxicity" pathways in conjunction with cell type specific functional readouts.

Generally, the obtained data should be compared across diverse *in vitro* models by extrapolation and further aligned to *in vivo* available data sets, in order to bridge the gap between *in vivo* and *in vitro* testing and prove the reliability of *in vitro* results.

Test Compounds

Applicability:

The iPSC-derived neuronal and glia cells have been used for testing CREB pathway perturbations elicited upon exposure to 25µM 2-naphthol-AS-E-phosphate (KG-501) (Pistollato et al., 2014). The CREB pathway is crucial in the development of the central nervous system, including neuronal, survival, neurite outgrowth, precursor proliferation and neuronal differentiation (Lonze et al., 2002).

KG-501 is a specific CREB signalling inhibitor (used as a positive control), which was administered in a time-course mode in the hiPSC-derived neuronal/glia cultures. The results have revealed that CREB pathway is expressed in IMR90-hiPSC-derived neurons and its inhibition is relevant to neurotoxicity testing. This cellular model might be a useful tool for studying chemical effects on human neuronal cells, providing a valuable alternative to the use of other *in vitro* animal models (Pistollato et al., 2014).

4. Discussion

Ethical issues and considerations for 3R's impact

The use of hiPSC cultures (derived from adult cells, such as fibroblasts) does not create as many ethical concerns as those linked to the use of human embryonic stem cells.

The presented method describes a robust and relatively rapid (28 days) protocol for the differentiation of hiPSCs (IMR90-hiPSCs) into heterogeneous cultures of post-mitotic neurons and glia.

Currently, the use of PSC-derived, and particularly, of hiPSC-derived cells for regulatory toxicity testing has been taken into great consideration (Kumar et al., 2012). Such test system offer several advantages, such as the capability to: (i) generate disparate and toxicologically relevant cell types, (ii) avoid the diverse national legislative policies regulating the use of human embryonic stem cell-derived models (hESCs), (iii) decrease/replace the use of animal and cancer-derived cell lines by human cell culture, and (iv) avoid complex extrapolation issues associated with use of animal models.

In this regard, heterogeneous cell derivatives, such as the hiPSC-derived neuronal and glial cultures, rather than homogenous cell cultures, might better mimic possible toxicological responses of human tissues *in vivo*.

Potential for future development

It would be beneficial to develop a working cell bank of neural stem cells (NSCs) and, when needed, further differentiate them into neurons and glia instead of direct differentiation of rosettes into mature neuronal cultures. This could further improve the robustness and reproducibility of the method. The procedure described in the **DB-ALM Protocol n° 166** was developed for that purpose and shows that the expansion and cryostorage of NSC's is feasible and the cells retain differentiation potential.

In order to demonstrate the value of neuronal/glia cell cultures derived from PSCs for toxicology, establishing and maintaining cell phenotypes over long-term cultures is of pivotal importance, as well as achieving automated cell production and differentiation. The development of mass cell production of hiPSC-derived neurons can be a challenging task, although mandatory for the final transfer of an assay on industrial platforms which could allow the automated screening for potential neurotoxicant.

Related methods

One of the aims of the EU Integrated Project SCR&Tox was to assess the activation of Keap1-Nrf-2-ARE pathway, which is related to oxidative stress, across different cell types in order to evaluate the relevance of this pathway as a general (horizontal) “toxicity” pathway in conjunction with cell type specific functional readouts, including neural cells. In order to address the multiple aspects of neurotoxicity, three different neuronal models have been established from different stem cell sources:

- a. hiPSC-derived neuroepithelial-like stem cells differentiated into motor neurons
- b. human embryonic stem cells differentiated into post-mitotic neurons
- c. mixed post-mitotic neuronal and glial cultures derived from hiPSCs (JRC's STU approach).

One of the advantages of the JRC method is that the hiPSCs can be used for regulatory toxicity testing without the diverse legislative policies that are required for the use of human embryonic stem cells. Another important aspect of this method is that the established differentiation protocol leads to heterogeneous cultures of post-mitotic neurons together with glial cells (see [DB-ALM Protocol No. 165](#)) which might better mimic toxicological responses of human tissues *in vivo*, than homogenous neuronal cultures (without glial cells).

One of the limitations of the DB-ALM Protocol No. 165 was the number of differentiated cells produced in one experimental run, which might be insufficient for large scale (high-throughput) chemical exposure studies. To overcome this limitation, an additional protocol has been developed by the JRC, which allows the expansion of rosette-derived NSCs, before further differentiation is induced (see [DB-ALM Protocol No. 166](#)).

5. Contact Details

Dr Anna Price

Chemicals Safety and Alternative Methods Unit - European Union Reference Laboratory for Alternatives to Animal Testing (EURL-ECVAM)
European Commission - Joint Research Centre
Corresponding author
via E. Fermi 1
Ispra (VA) 21027
email: Anna.Price@ec.europa.eu
telephone: 39 0332 7813527
last contact update: 30 May 2013

Dr Dimitra Zagoura

Chemicals Safety and Alternative Methods Unit - European Union Reference Laboratory for Alternatives to Animal Testing (EURL-ECVAM)
European Commission - Joint Research Centre
Method developer / user
via E. Fermi 1
Ispra (VA) 21027
telephone: 39 0332 7813527

Dr David Canovas-Jorda

Chemicals Safety and Alternative Methods Unit - European Union Reference Laboratory for Alternatives to Animal Testing (EURL-ECVAM)
European Commission - Joint Research Centre
Method developer / user
via E. Fermi 1
Ispra (VA) 21027
telephone: 39 0332 7813527

Dr Francesca Pistollato

Chemicals Safety and Alternative Methods Unit - European Union Reference Laboratory for Alternatives to Animal Testing (EURL-ECVAM)
European Commission - Joint Research Centre
Method developer / user
via E. Fermi 1

Ispra (VA) 21027
 email: Francesca.PISTOLLATO@ec.europa.eu
 telephone: 39 0332 7813527

Ver n°	Date	Name	Type of change
1	05 June 2015	Dr Dimitra Zagoura	Datasheet compiled

6. Abbreviations & Definitions

DB-ALM: EURL ECVAM Database Service on Alternative Methods to Animal Experimentation

EURL ECVAM: The European Union Reference Laboratory for alternatives to animal testing

hESC: human embryonic stem cells

hiPSC: human induced pluripotent stem cells

HCI: High Content Imaging

IC: Immunocytochemistry

JRC: Joint Research Center

MEA: Multielectrode Array

ND medium: Neural Differentiation medium

NI medium: Neural Induction medium

NSC: Neural Stem Cell

PSC: Pluripotent Stem Cell

Scr&Tox: EU Integrated project: "Stem cells for relevant efficient extended and normalized toxicology"

7. Bibliography

Method development:

- DB-ALM (2016)
 DB-ALM Protocol n° 166: Standard operating procedure for expansion of rosette-derived neural stem cells
[EURL ECVAM Database service on Alternative Methods to animal experimentation](#)
- DB-ALM (2016)
 DB-ALM Protocol n°165 : Standard operating procedure for differentiation of human induced pluripotent stem cells into post-mitotic neurons and glial cells (mixed culture)
[EURL ECVAM Database service on Alternative Methods to animal experimentation](#)
- Pistollato, F., Lousse, J., Scelfo, B., Mennecozzi, M., Accordi, B., Basso, G., Gaspar, J.A., Zagoura, D., Barilari, M., Palosaari, T., Sachinidis, A., Bremer-Hoffmann, S. (2014)
 Development of a pluripotent stem cell derived neuronal model to identify chemically induced pathway perturbations in relation to neurotoxicity: effects of CREB pathway inhibition.
Toxicology and Applied Pharmacology 280, 378–388

General references:

- Canovas-Jorda, D., Lousse, J., Pistollato, F., Zagoura, D., Bremer, S., (2014)
 Regenerative toxicology: the role of stem cells in the development of chronic toxicities.
Expert Opinion on Drug Metabolism & Toxicology 10(1), 39-50
- Coecke, S., Balls, M., Bowe, G., Davis, J., Gstraunthaler, G., Hartung, T., Hay, R., Merten, O.W., Price, A., Schechtman, L., Stacey, G. and Stokes, W. (2005)
 Guidance on Good Cell Culture Practice.
Alternatives to Laboratory Animals (ATLA) 33, 261-287

- Ghosh, A., Manrique-Hoyos, N., Voigt, A., Schulz, J.B., Kreutzfeldt, M., Merkler, D., Simons, M., (2011)
Targeted Ablation of Oligodendrocytes Triggers Axonal Damage.
PLoS One 6(7), e22735
- Grskovic, M., Javaherian, A., Strulovici, B., Daley, G.Q., 2011. Induced pluripotent stem cells — opportunities for disease modelling and drug discovery. (2011)
Induced pluripotent stem cells — opportunities for disease modelling and drug discovery.
Nature Reviews. Drug Discovery 10, 915–929
- Kumar, K.K., Aboud, A.A., Bowman, A.B. (2012)
The potential of induced pluripotent stem cells as a translational model for neurotoxicological risk.
Special Review.
Neurotoxicology 33, 518–529
- Lonze, B.E. & Ginty, D.D., (2002)
Function and Regulation of CREB Family Transcription Factors in the Nervous System.
Neuron 35(4), 605-623
- Pistollato, F., Bremer-Hoffmann, S., Healy, L., Young, L., and Stacey, G. (2012)
Standardization of pluripotent stem cell cultures for toxicity testing.
Expert Opinion on Drug Metabolism & Toxicology 8, 239–257
- Trosko, J.E. & Chang, C.-C., (2010)
Factors to consider in the use of stem cells for pharmaceutic drug development and for chemical safety assessment.
Toxicology 270(1), 18-34

Last update: 05 June 2015

DB-ALM Protocol n° 165 : Standard operating procedure for differentiation of human induced pluripotent stem cells into post-mitotic neurons and glial cells (mixed culture)

Stem cell culture method, Neurotoxicity, Systemic Toxicity

Detailed differentiation protocol for induced-pluripotent stem cells differentiated into neural precursors and further into mixed cultures of post-mitotic neurons and glial cells was developed within the SCR&Tox EU integrated project (*Seurat-1*).

Résumé

For background information on the scientific rationale and applications of this method please consult **DB-ALM Method Summary N°165: Differentiation of induced-pluripotent stem cells into post-mitotic neurons and glial cells (mixed culture)**.

Experimental Description

Experimental System(s):

hiPSs (IMR90-hiPSC culture) - Human induced-pluripotent stem cells (hiPSC) derived from human foetal lung fibroblast (IMR90), reprogrammed towards pluripotency by retroviral transduction of Oct4 and Sox2, using the pMIG vectors. Created and distributed by I-Stem (Evry, France).

Basic Procedure

The procedure takes approximately 28 days to complete, with main steps outlined in the **Figure 1** (see page 8). The hiPSC cells are seeded at appropriate density and medium and cultured for 10 days. The cells begin to form colonies which are cut and plated on low attached petri dishes where they form embryoid bodies (EBs) floating in the medium. EBs are plated on adherent matrix in the presence of the neuroepithelial induction medium and begin to create visible neuroepithelial aggregates (rosettes) after 5-8 days. Finally, rosettes are cut, dissociated and the cells plated onto an adherent surface of target vessel (MEA chip, 96-well plate or a Petri dish) and further cultured in the presence of neural differentiation medium for 21 days, where they generate mixed (neural/glial) post-mitotic neuronal cultures.

Status

Known Laboratory Use:

EURL ECVAM

Participation in Evaluation Studies:

The SOP was developed within the frame of the Scr&Tox EU Integrated project:

- <http://www.scrtox.eu/>

Proprietary and/or Confidentiality Issues

For this method the hiPS(IMR90 cell line), created and provided by I-Stem (Evry, France), was used as induced-pluripotent stem cells. On 8th October 2012 JRC has signed a material transfer agreement with I-Stem, as a partner of Scr&Tox project. The cells were used for non-profit research purposes.

Abbreviations and Definitions

bFGF: basic Fibroblast Growth Factor

BDNF: Brain-Derived Neurotrophic Factor

BSA: Bovine Serum Albumin

DB-ALM: EURL ECVAM Database Service on Alternative Methods to Animal Experimentation

DPBS: Dulbecco's Phosphate-Buffered Saline

EB: Embryoid body

EURL ECVAM: The European Union Reference Laboratory for alternatives to animal testing

GDNF: Glial-Derived Neurotrophic Factor

hiPSC: human induced Pluripotent Stem Cell

HCI: High Content Imaging

IC: Immunocytochemistry

MEA: Multielectrode Array

ND medium: Complete Neural Differentiation medium

NRI medium: Complete Neuroepithelial Induction Medium

NSC: Neural Stem Cell

P/S: Penicillin/Streptomycin

PSC: Pluripotent Stem Cell

qPCR: Quantitative Real-time PCR

Scr&Tox: EU Integrated project: *"Stem cells for relevant efficient extended and normalized toxicology"*

SOP: Standard Operating Procedure

Last update: 05 Jun 2015

PROCEDURE DETAILS, 05 June 2015

Standard operating procedure for differentiation of human induced pluripotent stem cells into post-mitotic neurons and glial cells (mixed culture) DB-ALM Protocol n° 165

Unless otherwise stated, all procedures must be carried out under aseptic conditions in a class II laminar flow cabinet (biological hazard standard).

Differentiated cells can be analysed using routine analytical techniques such as qPCR, IC, Affymetrix microarrays, HCI system or MEA analysis (these procedures are not included in the protocol).

Contact Details

Dr Anna Price

System Toxicology Unit - European Union Reference Laboratory for Alternatives to Animal Testing (EURL-ECVAM)
European Commission - Joint Research Centre
via E. Fermi 1
Ispra (VA) 21027
email: Anna.Price@ec.europa.eu
telephone: 39 0332 7813527

Dr Dimitra Zagoura

System Toxicology Unit - European Union Reference Laboratory for Alternatives to Animal Testing (EURL-ECVAM)
European Commission - Joint Research Centre
via E. Fermi 1
Ispra (VA) 21027
telephone: 39 0332 7813527

Dr David Canovas-Jorda

System Toxicology Unit - European Union Reference Laboratory for Alternatives to Animal Testing (EURL-ECVAM)
European Commission - Joint Research Centre
via E. Fermi 1
Ispra (VA) 21027
telephone: 39 0332 7813527

Dr Francesca Pistollato

System Toxicology Unit - European Union Reference Laboratory for Alternatives to Animal Testing (EURL-ECVAM)
European Commission - Joint Research Centre
via E. Fermi 1
Ispra (VA) 21027
email: Francesca.PISTOLLATO@ec.europa.eu
telephone: 39 0332 7813527

Materials and Preparations

Cell or Test System

hiPSC line: wild type IMR90 (received from I-Stem)

Equipment

Fixed Equipment

- Benchtop centrifuge
- Countess® Automated Cell Counter
- Fridge (4°C)
- Freezer (-20°C)
- HCI using Cellomics platform
- Humidified incubator (37°C, 5% CO₂ in air)
- Laminar flow hood for sterile atmosphere (type II classified)

- qPCR Sequence Detection System
- Optical inverted microscope
- Stereoscopic (dissecting) microscope
- Micropipettes
- Water Bath

Consumables

(Cell culture, plastic ware and disposables)

- T-25 or T-75: 25 and 75 cm² cell culture flasks with filtered cap (Corning, Cat No. 430639 and TPP, Cat No. 90076, respectively)
- 15 and 50 ml conical plastic tubes (TPP, Cat No. 91015 and 91050, respectively)
- 5 and 10 ml plastic pipettes (Dispenser (TPP), Cat No. 94005 and 94010, respectively)
- 96-well plates (TPP, Cat No. 92096)
-
- Countess Cell Counting Chamber (Invitrogen, Cat No. C10283)
- Eppendorf tubes
- Gibson 20 and 300 µl dualfilter pipette tips (Eppendorf, Cat No. 022491270 and 022491245, respectively)
- Gibson 20-200 µl filter pipette tips (Diamond, Cat No. GF-F171503)
- Low attachment 60 mm petri dishes (Greiner, Cat No. 628102)
- MEA chips (single well, Multichannel Systems)
- Powder-free Nitrile Gloves (Microflex)

Media, Reagents, Sera, others

- 30G needle
- B27 Serum-Free Supplement (Invitrogen, Cat No. 17504044)
- BSA (7.5% solution) (Sigma, Cat No. A8412)
- BDNF (Invitrogen, Cat No. PHC7074)
- bFGF (Invitrogen, Cat No. 13256-029 or Immunological Sciences, Cat No. GRF-15595)
- β-Mercaptoethanol (50 mM) (ThermoFisher, Cat No. 31350-010)
- DMEM/F12 (Invitrogen, Cat. No. 3133-038)
- DPBS no calcium, no magnesium (1X) (Gibco, Cat No. 14190-094)
- EGF Recombinant Human Protein, Animal-Origin Free (Invitrogen, PHG6045)
- GDNF (Invitrogen, Cat No. PHC7045)
- Heparin Grade I-A, ≥180 USP units/mg (Sigma, Cat No. H3149-100KU)
- L-Glutamin 200 mM Solution (Gibco, Cat No. 25030-081)
- Laminin (Sigma, Cat No. L2020)
- Knockout DMEM (Invitrogen, Cat No. 10829-018)
- Knockout Serum Replacement (Invitrogen, Cat No. 10828-028)
- Matrigel™ hESC-qualified Matrix (Corning, Cat No. 354277)
- Matrigel® Basement Membrane Matrix (Corning, Cat No. 354234)
- N2 Supplement (Invitrogen, Cat No. 17502-048)
- Neurobasal Medium™ (Invitrogen, Cat No. 21103049)
- NEAA: Non-Essential Amino Acids (Invitrogen, Cat No. 11140-035)
- P/S: Penicillin/Streptomycin (Invitrogen, Cat No. 15140-122)
- Poly-L-ornithine 0.01% solution, 50 ml (Sigma, Cat No. P4957)
- Sterile-Filtered Water (Sigma, Cat No. W3500)
- Trypan blue (Sigma, Cat. No. T8154), for manual cell counting
- mTeSR™1 Basal Medium, (Stem Cell Technologies, Cat No. 05851)
- mTeSR™1 5X Supplements, (Stem Cell Technologies, Cat No. 05852)

Preparations

Media and Endpoint Assay Solutions

1. *BDNF Aliquots*

- Dissolve 10 µg of BDNF in 1 ml of 0.1% BSA/DPBS
- Aliquot in 50 µl (10µg/ml) and store at -20°C.

2. *bFGF Aliquots*

Add 10 ml of 0.1 % BSA-DPBS to 100 µg of bFGF for a final concentration of 10 µg /ml

- Aliquot bFGF solution in 500 µl and store at -20°C.

3. *0.1 % BSA/DPBS*

- Add 1.33 ml BSA (7.5%) to 98.67 ml of DPBS for a final BSA concentration of 0.1%

4. *EGF Aliquots*

- Dissolve 500 µg of human EGF in 5 ml of 0.1% BSA/DPBS and aliquot in 1 ml (100 µg/ml) store at -20°C.
- Add 9 ml of 0.1% BSA/DPBS to 1 ml (100 µg/ml) for a final concentration of 10 µg/ml.
- Aliquots in 500 µl and store at -20°C

5. *L-Glutamin 200 mM*

- Thaw out L-glutamin bottle
- Prepare 5 ml aliquots and store at -20°C

6. *Heparin*

- Dissolve heparin in DMEM/F12 medium (1mg/ml)
- Aliquot solution (1ml aliquots) and store at -20 °C

7. *Laminin Coating Solution (optional)*

Laminin coating solution is freshly prepared before use

- Thaw out laminin
- Dilute neat laminin (1mg/ml) in DPBS for a final laminin concentration of 10 µg/ml (1:100 dilution)

8. *Matrigel™ Basement Membrane Matrix/ Matrigel™ hESC-Qualified Matrix (recommended)*

- Thaw out the Matrigel™ Basement Membrane Matrix or Matrigel™ hESC-Qualified Matrix by placing the matrix vial/bottle on ice until matrix becomes liquid (do not let the matrix to solidify at temperatures above 10°C by keeping it on ice)
- Aliquot matrix in 200 µl and store them at -20°C until use

9. *Penicillin/Streptomycin Aliquots*

- Thaw out P/S bottle
- Prepare 2.5 ml aliquots and store at -20°C

10. *GDNF Aliquots*

- Dissolve 10 µg of GDNF in 1 ml of 0.1% BSA/DPBS, aliquot in 50 µl (10 µg/ml) and store at -20°C.

11. *Cell Media Preparations*

Note: Some media recipes below include antibiotic Penicillin/Streptomycin mix which is optional. When hiPSC-derived neuronal cultures will be used particularly for analysis of electrical activity, the use of P/S mixture should be avoided if possible. It alters the electrophysiological properties of cultured neurons and may interfere with the MEA analysis (Bahrami and Janahmadi, 2013).

• **Complete Feeder-Independent Maintenance Medium (for hiPSC expansion)**
(Complete hiPSC medium)

	Volume 50 ml	Volume 500 ml ⁽¹⁾
mTeSR™1 Basal Medium	40 ml	400 ml
	10 ml	100 ml

(1) Complete mTeSR™1 is stable when stored at 2 - 8°C for up to 2 weeks or when frozen at -20°C for up to 6 months. Thaw frozen complete medium at room temperature (15 - 25°C) or overnight at 2 - 8°C before use.

(2) 5X Supplements can be dispensed into working aliquots and stored at -20°C. Use frozen aliquots within 3 months. Thawed aliquots should be used within 1 day to prepare complete mTeSR™1 medium.

• **Complete iPSC EB Medium (for EB formation)**

	Final Concentration	Volume 50 ml	Volume 250 ml
Knockout DMEM	1 X	39 ml	195 ml
Knockout Serum Replacement	20%	10 ml	50 ml
NEAA	1 X	500 µl	2.5 ml
Penicillin/Streptomycin	50 U/ml	250 µl	1.25 ml
L-Glutamine	2 mM	500 µl	2.5 ml
β-Mercaptoethanol	50 µM	50 µl	250µl

• **Complete Neuroepithelial Induction (NRI) Medium (for neuroepithelial cell formation)**

	Final Concentration	Volume 50 ml	Volume 500 ml
DMEM/F12	1 X	49 ml	490 ml
NEAA	1 X	500 µl	5 ml
N2 Supplement	1 X	500 µl	5 ml
Penicillin/Streptomycin	50 U/ml	250 µl	2.5 ml
Heparin	2 µg/ml	100 µl	1 ml
bFGF	20 ng/ml	100 µl	1000 µl ⁽³⁾

(3) bFGF is instable therefore complete NRI medium should be used within 2-3 weeks.

• **Complete Neural Differentiation (ND) Medium**

	Final Concentration	Volume 50 ml	Volume 500 ml
Neurobasal Medium™	1 X	50 ml	500 ml
B-27 Supplements (50x)	1 X	1 ml	10 ml
GDNF	1 ng/ml	5 µl	50 µl
BDNF	2.5 ng/ml	12.5 µl	125 µl
Penicillin/Streptomycin	50 U/ml	250 µl	2.5 ml

Method

Routine Culture Procedure

A. *Matrigel™ Basement Membrane Matrix and Matrigel™ hESC-qualified matrix Coating of Culture Dishes**

- *Matrigel™ basement membrane matrix* is used for expansion of neural stem cells; *Matrigel™ hESC-qualified matrix* is used for expansion of pluripotent stem cells
- Thaw out 200 µl of the appropriate Matrigel™ aliquots (see point 8. in *Media and Endpoint Assay Solutions* above) on ice
- Dilute 200 µl of Matrigel™ in 20 ml of DMEM/F12 medium
- Add 5 ml of *Matrigel™ /DMEM/F12* solution to a 60 mm Petri dish and leave it in an incubator at 37 °C for a minimum of 1 hour
- Remove *Matrigel™ /DMEM/F12* solution before cell plating

* Note : Alternative to the use of *Matrigel™ basement membrane matrix*, *Laminin* or *Poly-L-ornithine/Laminin* coating can be used. These procedures take longer and the cell attachment is less strong, however they might be e.g. less likely to interfere with the electrical reading on MEA chips

Laminin Coating

- Add 4 ml (or 8 ml) of laminin coating solution (see point 7. in *Media and Endpoint Assay Solutions* above) to a T-25 (or T-75) flask
- Place flask in an incubator at 37 °C for a minimum of 2 hours
- Ensure all laminin solution is removed from the flask before adding cells

Poly-L-ornithine/Laminin Coating

- Dilute Poly-L-ornithine (1:6) in DPBS to coat plates and flasks (e.g., 100 µl/well in 96 well plate)
- Incubate flasks/plates overnight (or minimum 6 hours) at 37°C
- Rinse 3 times with DPBS
- Coat with 1:500 laminin coating solution (see point 7. in *Media and Endpoint Assay Solutions* above) for minimum 2 hours
- Aspirate laminin before plating the cells

B. *Cell Counting (Trypan blue dye)*

- Under sterile conditions take 11 µl of your cell suspension into a sterile Eppendorf tube
- Add 11 µl of Trypan Blue dye to your 11 µl of cell suspension (1:1), mix and place 11 µl in a Countess cell counting chamber slide
- Place slide in the Countess® Automated Cell Counter to obtain the number of cells per ml

C. *Pluripotent Stem Cell Culture and Differentiation*

1. *hiPSC Cell Culture Maintenance*

hiPSC (wild type IMR90) cultures are grown on *Matrigel™ hESC-qualified matrix* -coated 60 mm Petri dishes

- Coat a 60 mm Petri dish with *Matrigel™ hESC-qualified matrix* minimum one hour before passaging cells
- Using a phase contrast or a stereoscopic (dissecting) microscope under sterile conditions (in a laminar flow cabinet) cut stem cell colonies in squares of about 200µm x 200µm using a 30G needle
- Aspirate in and out growth medium using a 200µl pipette to detach colony pieces.
- Aspirate detached colony pieces (around 100 pieces) and place them in a fresh *Matrigel™ hESC-qualified matrix* -coated Petri dish. Add 3-4 ml of **complete hiPSC medium**
- Place Petri dish in an incubator at 37°C, 5% CO₂
- Perform total medium change every day

2. Differentiation protocol for post-mitotic neurons from hiPSC cultures

The neuronal differentiation protocol applied on IMR90 cells has been developed based on the procedures used by Stummann et al. (2009) and Pistollato et al. (2014) with some modifications (**Figure 1**).

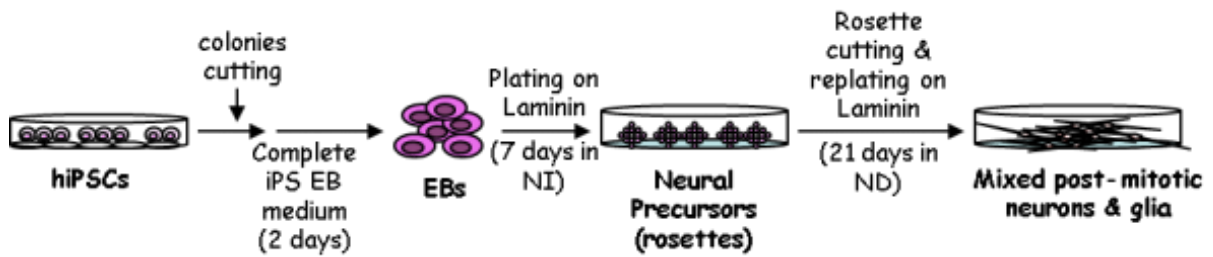


Figure 1: Cartoon summarizing the differentiation protocol for PSC-derived post-mitotic neurons used at the JRC.

a. Formation of PSC aggregates (EBs) (day 0-day 2)

Day 0:

- Change **complete hiPSC medium** (with 3ml of medium/60 mm Petri dish) prior cutting the undifferentiated colonies under sterile conditions
- Cut colonies with 30G needle into fragments of approximately 200µm x 200µm
Note: This step requires great manual skill and precision and is critical for the procedure. The iPSC colony fragments must be of equal size to obtain homogenous EBs in the next step
- Detach fragments very gently using a p200 µl tip
- Transfer detached fragments into a 15 ml tubes, rinsing the dish with additional 2 ml of **complete hiPSC medium**
- Centrifuge at 800 rpm (112 g) for 1 min
- Aspirate supernatant and gently resuspend pellet in 5 ml of **complete iPS EB medium**
- Place cell suspension in 60 mm low attachment Petri dish (5 ml per dish)
- Leave Petri dish overnight in an incubator at 37°C, 5% CO₂
Note: Normally colonies coming from 1x60 mm Petri dish can be re-plated as floating aggregates in 1x60mm low-attachment Petri dish.

Day 1:

Note: floating EBs should be already visible

- Transfer EBs into a 15 ml tube
- Centrifuge EBs at 800 rpm (112 g) for 1 min
- Remove cell debris by decanting supernatant (be careful not to decant the EBs)
- Resuspend EBs in 5 ml **complete iPS EB medium** and transfer them into 60 mm low attachment Petri dish
- Return dish to the incubator at 37°C, 5% CO₂
- Coat 1x60 mm petri-dishes (1 for every 60mm petri dishes containing EBs) with *Matrigel™ Basement Membrane Matrix* or laminin (to be used the day after, day 2)

b. *Early neural differentiation: neuroepithelial aggregates (rosettes) formation (day 2-day 7)*

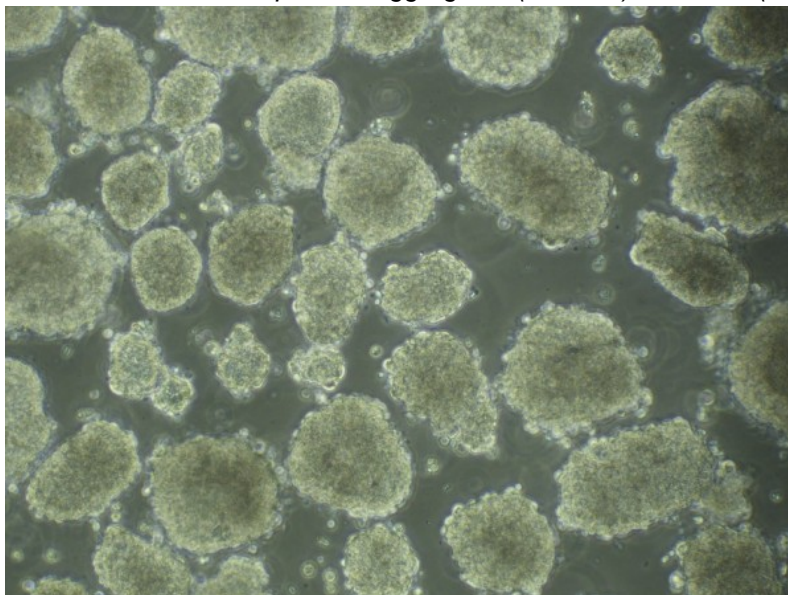


Figure 2: Phase contrast image of IMR90 embryoid bodies before plating (10x magnification).

Day 2:

- Remove laminin or Matrigel coating solution from 60 mm Petri dishes and plate EBs in 4 ml of **complete NRI medium** (about 50 EBs/60 mm in Petri dishes)
- Perform a total medium change with **complete NRI medium** every other day, up to the 7th day
- After maximum 5 days (i.e. 7th day) neuroepithelial aggregates (rosettes) should be visible.

-
- Coat with *Matrigel™ Basement Membrane Matrix* or laminin at 37°C (or for a minimum of 2 hours before use) any required 96 well plates (for IC, 100µl/well), 24 well plates (for qPCR, 250µl/well), 12 well plates (for Affymetrix, 500µl/well), MEA chips (for electrical activity, 1ml/chip), or 60 mm Petri dishes (for NSC expansion)

c. *Late neuronal differentiation:*

Cutting of rosettes like structure and post mitotic neurons differentiation (Day 8-Day 28)

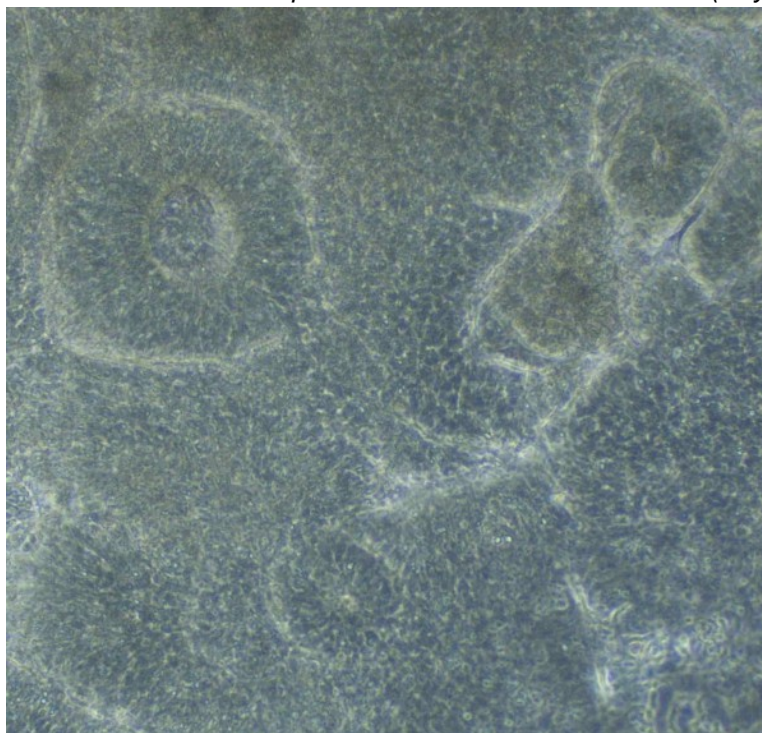


Figure 3: Phase-contrast image of IMR90 rosettes at Day 8, visible as cylindrical structures (10x magnification).

Day 8:

Using a phase contrast or a stereoscopic (dissecting) microscope under sterile conditions, cut rosettes-like structures in very small fragments from 60 mm Petri dishes using a 30G needle and gently perform fragment detachment by pipetting up and down medium using a p100 tip

- Transfer rosettes-like structures fragments into a 15 ml conical tube and spin down at 800 rpm (112 g) for 2 min
- Remove supernatant and gently resuspend pellet in 1 ml of DPBS (to perform a partial single cell dissociation)
- Add appropriate volume of **complete NRI medium**
- Count cells with trypan blue (see section B. **Cell Counting (Trypan blue dye)** above, this step may be omitted if cells cannot be brought into a single cell suspension)
- Aspirate the laminin from Petri dishes, 96 well plates and/or MEA chips.
- Plate single/partial dissociated cells in **complete NRI medium** according to the study plan. For example in some JRC studies a density of approximately 15.000 cells/cm² was used.

Note: If rosette fragments do not look completely dissociated, in order to reach a cell plating density of about 15.000 cells/cm², dissociated rosettes fragments deriving from about 50 EBs/1x60mm Petri dish can be resuspended in 50 ml of complete NRI medium and plated as follows, depending on the selected plate format:

Multiwell plates/MEA	Growth area (cm ²)	Volume of cell suspension to plate per well (or MEA chip)	Maximum number of plates that can be plated (with 50 ml of cell suspension)
96 wells	0.3	100 µl	5
48 wells	0.7	220 µl	4
24 wells	2	625 µl	3
12 wells	4	1.25 ml	3
6 wells	10	3.125 ml	2
Single well MEA chip	3.5	1.1 ml	45

Additional considerations:

- i. The selected EBs should be homogenous in size if iPSC colony fragments of approximately 200µm x 200µm have been generated in step *a. Formation of PSC aggregates (EBs) Day 0*
- ii. At this stage the cells can be either further differentiated into a mixed culture of the post-mitotic neurons and glial cells, or expanded and maintained in the present state, to be fully differentiated at a later stage.
The latter procedure is described in the **DB-ALM Protocol No. 166: Standard operating procedure for expansion of rosette-derived neural stem cells**

Day 10:

To promote full differentiation of rosette derived cells:

- After two days in **complete NRI medium**, perform a medium change using **complete ND medium**

Day 12 onwards:

- Perform total medium change with **complete ND medium** every 3-4 days (twice a week) until day 28
- On day 28 process the cells according to the planned readout (using eg. RNAqueous®-Micro Kit for RNA extraction for qPCR).

DB-ALM Protocol n° 166 : Standard operating procedure for expansion of rosette-derived neural stem cells

Stem cell culture method, Neurotoxicity, Systemic Toxicity

Procedure for expansion of rosette-derived neural precursor stem cells, intended for further differentiation into mixed cultures of post-mitotic neurons and glial cells. Technique developed within the SCR&Tox EU integrated project (*Seurat-1*).

Résumé

For background information please consult **DB-ALM Method Summary N°165: Differentiation of induced-pluripotent stem cells into post-mitotic neurons and glial cells (mixed culture)**.

Experimental Description

Experimental System(s):

hiPSs (IMR90-hiPSC culture) - Human induced-pluripotent stem cells (hiPSC) derived from human foetal lung fibroblast (IMR90), reprogrammed towards pluripotency by retroviral transduction of Oct4 and Sox2, using the pMIG vectors. Created and distributed by I-Stem (Evry, France).

Basic Procedure

The rosette-derived neural stem cells (NSC) are prepared from the neuroepithelial aggregates (rosettes), obtained as described in the **DB-ALM Protocol N°165** (*step c: Late neuronal differentiation: Cutting of rosettes like structure on Day 8*) (DB-ALM, 2016).

The NSC's are seeded at the density of 50,000 cell/cm² in Complete Neural Induction (NI) Medium and grown until confluence. Afterwards NSC's are trypsinised, counted, diluted to a desired density and plated into fresh medium.

Discussion

This procedure was developed as an extension to the **DB-ALM Protocol N°165**. It can be used to maintain and expand the rosette-derived NSCs (obtained from hiPS(IMR90) cell line), in an undifferentiated, proliferative state. In this way an increased number of cells can be obtained for further differentiation and chemical exposure studies.

Status

Participation in Evaluation Studies:

The SOP was developed within the frame of the Scr&Tox EU Integrated project:

- <http://www.scrtox.eu/>

Proprietary and/or Confidentiality Issues

This procedure is intended for the experiments conducted with the hiPS(IMR90 cell line), created and provided by I-Stem (Evry, France) and used as iPSCs. On 8th October 2012 JRC has signed a material transfer agreement with I-Stem, as a partner of Scr&Tox project. The cells were used for non-profit research purposes.

Abbreviations and Definitions

bFGF: basic Fibroblast Growth Factor 2

BDNF: Brain-Derived Neurotrophic Factor

BSA: Bovine Serum Albumin

DB-ALM: EURL ECVAM Database Service on Alternative Methods to Animal Experimentation

DPBS: Dulbeccos Phosphate-Buffered Saline

EURL ECVAM: The European Union Reference Laboratory for alternatives to animal testing

hiPSC: human induced Pluripotent Stem Cell

IC: Immunocytochemistry

NI medium: Complete Neural Induction medium

NRI medium: Complete Neuroepithelial Induction medium

NSC: Neural Stem Cell

P/S: Penicillin/Streptomycin

Scr&Tox: EU Integrated project: "*Stem cells for relevant efficient extended and normalized toxicology*"

SOP: Standard Operating Procedure

T-25: 25 cm² cell culture flasks

Last update: 26-02-2016

PROCEDURE DETAILS, 26 November 2015

Standard operating procedure for expansion of rosette-derived neural stem cells DB-ALM Protocol n° 166

This procedure is complementary to the **DB-ALM Protocol N°165: Standard operating procedure for differentiation of human induced pluripotent stem cells into post-mitotic neurons and glial cells (mixed culture)**.

Unless otherwise stated, all procedures must be carried out under aseptic conditions in a class II laminar flow cabinet (biological hazard standard).

Contact Details

Dr David Canovas-Jorda

System Toxicology Unit - European Union Reference Laboratory for Alternatives to Animal Testing (EURL-ECVAM)
European Commission - Joint Research Centre
via E. Fermi 1
Ispra (VA) 21027
telephone: 39 0332 7813527

Dr Anna Price

System Toxicology Unit - European Union Reference Laboratory for Alternatives to Animal Testing (EURL-ECVAM)
European Commission - Joint Research Centre
via E. Fermi 1
Ispra (VA) 21027
email: Anna.Price@ec.europa.eu
telephone: 39 0332 7813527

Materials and Preparations

Cell or Test System

The rosette-derived neural stem cells (NSC) are prepared from the neuroepithelial aggregates (rosettes), obtained from the hiPS (IMR90 cell line), following the procedure described in the **DB-ALM Protocol N°165**, till *step c: Late neuronal differentiation: Cutting of rosettes like structure and post mitotic neurons differentiation (day 8)* (DB-ALM, 2016).

Equipment

Fixed Equipment

- Benchtop centrifuge
- Countess® Automated Cell Counter
- Fridge (4°C)
-
- Humidified incubator (37°C, 5% CO₂ in air)
- Ice machine
- Laminar flow hood for sterile atmosphere (type II classified)
- Liquid nitrogen storage
- Micropipettes
- Milli-Q water dispenser
- Mr. Frosty™ freezing container
- pH meter
- Stereoscopic (dissecting) microscope
- Water Bath

Consumables

(Cell culture, plastic ware and disposables)

- 15 and 50 ml conical plastic tubes (TPP, Cat No. 91015 and 91050, respectively)
- 25 cm² cell culture flasks with filtered cap (T-25) (Corning, Cat No. 430639)
- 5 and 10 ml plastic pipettes (Dispenser) (TPP, Cat No. 94005 and 94010, respectively)
- 60 mm Petri dishes (TPP, Cat No. 93060)
- Countess Cell Counting Chamber (Invitrogen, Cat No. C10283)
- Cryovials Nalgene® (Sigma Cat No. V5007-500EA)
-
- Gibson 20 and 300 µl dualfilter pipette tips (Eppendorf, Cat No. 022491270 and 022491245, respectively)
- Gibson 20-200 µl filter pipette tips (Diamond, Cat No. GF-F171503)
- Powder-free Nitrile Gloves (Microflex)

Media, Reagents, Sera, others

- B27 Serum-Free Supplement (Invitrogen, Cat No. 17504044)
- BDNF (Invitrogen, Cat No. PHC7074)
- BSA (7.5% solution) (Sigma, Cat No. A8412)
- bFGF (Immunological Sciences, Cat No. GRF-15595)
- Defined Trypsin Inhibitor (Invitrogen, Cat No. R-007-100)
- DMEM/F12 (Invitrogen, Cat. No. 3133-038)
- DMSO
- EGF Recombinant Human Protein, Animal-Origin Free (Invitrogen, PHG6045)
- Heparin Grade I-A, ≥180 USP units/mg (Sigma, Cat No. H3149-100KU)
- L-Glutamin 200 mM Solution (Gibco, Cat No. 25030-081)
- Matrigel® Basement Membrane Matrix (Corning, Cat No. 354234)
- N2 Supplement (Invitrogen, Cat No. 17502-048)
- NEAA: Non-Essential Amino Acids (Invitrogen, Cat No. 11140-035)
- 2-Propanol (Sigma, Cat No. 190764)
- P/S: Penicillin/Streptomycin (Invitrogen, Cat No. 15140-122)
- Sterile-Filtered Water (Sigma, Cat No. W3500)
- Trypan blue (Sigma, Cat No. T8154)
- Trypsin-EDTA (0.5%), no phenol red (Gibco, Cat No. 15400-054)
- mTeSR™ 1 Basal Medium, (Stem Cell Technologies, Cat No. 05851)
- mTeSR™ 1 5X Supplements, (Stem Cell Technologies, Cat No. 05852)

Preparations

Media and Endpoint Assay Solutions

1. *bFGF Aliquots*

- Add 10 ml of 0.1% BSA/DPBS to 100 µg of bFGF for a final concentration of 10 µg/ml
- Aliquot bFGF solution in 500 µl aliquots and store at 4°C.

2. *0.1 % BSA/ DPBS*

- Add 1.33 ml BSA (7.5%) to 98.67 ml of DPBS for a final BSA concentration of 0.1%

3. *BDNF Aliquots*

- Dissolve 10 µg of BDNF in 1 ml of 0.1% BSA/DPBS, aliquot in 50 µl (10µg/ml) and store at -20°C.

4. *EGF Aliquots*

- Dissolve 500 µg of human EGF in 5 ml of sterile water.
- Aliquot in 1 ml (100 µg/ml) and store at -20°C.

- Add 9 ml of 0.1% BSA/DPBS to 1 ml (100 µg/ml) for a final concentration of 10 µg/ml.
- Aliquot in 500 µl and store at -20°C .

5. Heparin

- Dissolve heparin in DMEM/F12 medium (1mg/ml)
- Aliquot solution (1ml aliquots) and store at -20 °C

6. L-Glutamin 200 mM

Thaw out L-glutamin bottle

- Prepare 5 ml aliquots and store at -20°C

7. Matrigel® Basement Membrane Matrix

- Thaw out Matrigel by placing Matrigel vial/bottle on ice until Matrigel becomes liquid (NOTE: do not let Matrigel to solidify at temperatures above 10°C by keeping it on ice)
- Aliquot Matrigel in 200 µl portions and store at -20°C until use

8. Penicillin/Streptomycin Aliquots

- Thaw out P/S bottle
- Prepare 2.5 ml aliquots and store at -20°C

9. Trypsin-EDTA (0.5%)

- Trypsin is diluted in DPBS to a final concentration of 0.05% (1:10 dilution)

10. Cell Medium Preparation

• **Complete Neural Induction (NI) Medium**

	Final Concentration	Volume 50 ml	Volume 500 ml
DMEM/F12	1 X	49 ml	490 ml
NEAA	1 X	500 µl	5 ml
N2 Supplement	1 X	500 µl	5 ml
Penicillin/Streptomycin	50 U/ml	250 µl	2.5 ml
Heparin	2 µg/ml	100 µl	1 ml
bFGF (1)	10 ng/ml	50 µl	500 µl
L-Glutamine	2 mM	500 µl	5 ml
EGF (1)	10 ng/ml	50 µl	500 µl
BDNF (1)	2.5 ng/ml	12.5 µl	125 µl
B-27 Supplements (50x)	1X	1 ml	10 ml

(1) bFGF is unstable and therefore the complete NI medium should be used within 2-3 weeks.

Alternatively, NI medium containing all other medium supplements except bFGF, EGF and BDNF can be prepared and divided in 50 ml aliquots to which bFGF, EGF and BDNF are added just before use.

• **Freezing medium**

	Final Concentration	Volume 10 ml
mTeSR™ 1 Basal Medium containing mTeSR™ 1 supplements and 50mg/ml BSA	50%	5 ml
Complete NI medium	40%	4 ml
DMSO	10%	1 ml

Method

Test System Procurement

Follow the procedure as described in the **DB-ALM Protocol N°165** until *step c*, Day 8: *Cutting of rosettes like structures* (DB-ALM, 2016).

Routine Culture Procedure

Matrigel® Basement Membrane Matrix Coating of Culture Dishes

1. Thaw out 200 µl Matrigel on ice
2. Dilute 200 µl of Matrigel in 20 ml of DMEM/F12 medium
3. Add 5 ml of *Matrigel DMEM/F12* solution to a 60 mm Petri dish/T-25 and leave it in an incubator at 37 °C for a minimum of 1 hour
4. Remove solution before cell plating

Neural Stem Cell Expansion

The neuronal stem cell expansion protocol applied on IMR90 cells is based on the protocol supplied by the cell provider (I-Stem, Evry (France)) with slightly modifications.

1. Spin down rosettes-like structures fragments in a conical 15ml tube at 800 rpm (112 g) for 2 min and gently resuspend the pellet in **NI medium**
2. Gently dissociate fragments by pipetting and transfer cells onto a Matrigel™ basement membrane matrix pre-coated 60 mm Petri dish/T-25 flask
3. Culture rosette-derived neural stem cells in **NI medium** performing medium change every other day until they reach confluence, as illustrated in **Figure 1**.

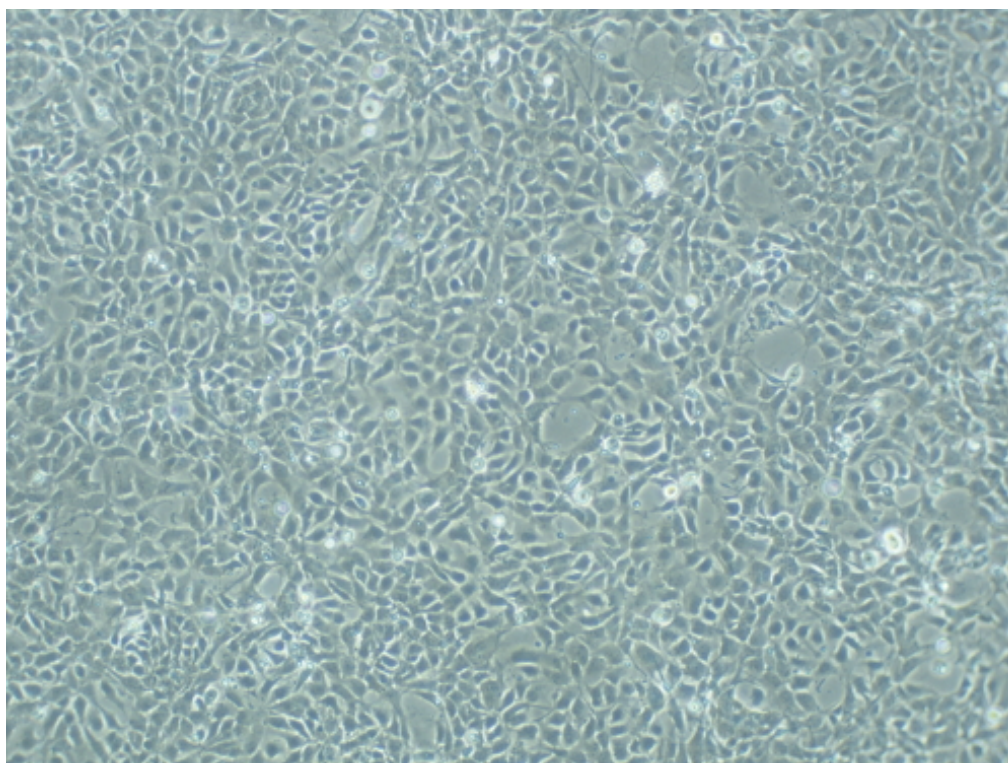


Figure 1: Phase-contrast image of confluent IMR90 neural stem cells (10x magnification) (10x magnification).

Neural Stem Cell Passage

1. Remove medium
2. Wash cells gently with warmed *DPBS*
3. Add 1.2-1.5 ml of warmed 0.05% *Trypsin-EDTA* to the dish containing the cells and place it in the

incubator for 1 minute

4. Check cell under the microscope and help cells to detach by gently tapping the flask
5. Add 1.2-1.5 ml of *Defined Trypsin Inhibitor*. Transfer cells into a 15 ml tubes
6. Wash Petri dish/T-25 flask by adding equal **NI medium** volume (i.e. 1.2-1.5 ml) and place it in the 15 ml tube containing the cells
7. Centrifuge the cells for 3 min at 130 g
8. Ensure that all Matrigel solution is removed from the new flask. Avoid Matrigel drying by adding a few ml of **NI medium** to wet the bottom of the flask.
9. Remove supernatant leaving around 100-200 µl to avoid removing the cells pellet
10. Dissociate cell pellet by resuspending cells in **NI medium** (e.g 500 µl) and gently mixing cell suspension
11. Add appropriate volume of **NI medium** to cell suspension (e.g. 3.5 ml)
12. Perform cell counting, see step **Cell Counting (Trypan blue dye)** below
13. Plate cells at density of about 50,000 cells/cm². During the first passages (1-4) it is recommended to seed cells at densities of at least 75,000 cells/cm²
14. Perform a total medium change every other day

Cell Counting (Trypan Blue dye)

1. Under sterile conditions take 11 µl of your cell suspension into a sterile Eppendorf tube
2. Add 11 µl of *Trypan Blue dye* to your 11 µl of cell suspension (1:1), mix and place 11 µl in a Countess cell counting chamber slide
3. Place slide in the Countess® Automated Cell Counter to obtain the number of cells per ml

Cyropreservation of NCS

1. Harvest the cells at point 7. of the **Neural Stem Cell Passage** procedure
2. Resuspend the cells in **Freezing Medium**, to achieve a cell concentration of 3x10⁶ cells/ml
3. Aliquot 0.5-0.53 ml of cell suspension, containing 1.5-1.6 x10⁶ cells per cryovial. The higher cell number will reach confluence faster when defrosted.
4. Place cryovials in *Mr. Frosty™ Freezing Container* (containing 2-propanol to achieve a rate of cooling very close to -1°C/minute)
5. Place *Mr. Frosty™ Freezing Container* at -80°C for minimum 2 hours and up to 2 weeks
6. Rapidly transfer cryovials into vapour phase of liquid nitrogen

Thawing of Cryopreserved NCS

1. Prepare one *Matrigel® Basement Membrane Matrix*-coated Petri dish to thaw one cryovial
2. Retrieve cryovial from the vapour phase of the liquid nitrogen tank and place immediately in warm water bath (37°C) for about 30-60 seconds
3. Transfer thawed cell solution in a 15 ml conical tube containing 6 ml of *warmed NI medium*
4. Centrifuge cells to remove DMSO for 4 min at 850 rpm
5. Carefully remove supernatant and very gently resuspend the cells in 1 ml **NI medium** (using 1 ml pipette tip)
6. Add 4 ml of **NI medium** into the 60 mm-Petri dish and carefully add 1 ml of cell suspension on top
7. Culture cells in the incubator
8. Refresh medium after 48 hours
9. When cells reach confluence, split them following the **Neural Stem Cell Passage** procedure

Bibliography


- DB-ALM (2016)
DB-ALM Method Summary n°165: Differentiation of induced-pluripotent stem cells into post-mitotic neurons and glial cells (mixed culture)
[EURL ECVAM Database service on Alternative Methods to animal experimentation](#)
- DB-ALM (2016)
DB-ALM Protocol n°165 : Standard operating procedure for differentiation of human induced pluripotent stem cells into post-mitotic neurons and glial cells (mixed culture)
[EURL ECVAM Database service on Alternative Methods to animal experimentation](#)



DNA methylation analysis using HumanMethylation450 microarray

Epigenomic data on the DNA methylation can be obtained, assessed and comparatively evaluated with the untreated controls using Illumina 450K BeadChip Array. It covers up to 485.000 CpGs spread over the genome (mostly in promoters and gene bodies). The method is cheap (~250 €/sample) and easy to perform. One of the known limitations is the the limited coverage of CpGs (repetitive elements are missing). This is one of the methods applied in the Large Scale Multi-omics VPA Case Study investigated by the NOTOX project. The researchers provided a detailed protocol for sample preparation when conducting epigenomic evaluations for all scientists interested in rough epigenetic profiling.

Method Summary n° 170.....55

Protocol n° 170..........61

Analysis of histone modifications via chromatin immunoprecipitation and sequencing (ChIP-seq)

ChIP-seq is the current state of the art regarding histone modification profiling, as it provides genome-wide data. There is a wide range of commercially available antibodies validated for ChIP-seq allowing a deep characterization of histone modification landscapes. This is one of the methods applied in the Large Scale Multi-omics VPA Case Study investigated by the NOTOX project. The genome-wide histone modifications of treated and control samples were studied using chromatin immunoprecipitation followed by deep sequencing. ChIP-seq data yield valuable information for the characterization of regulatory genomic regions.

Protocol n° 171.....67

DB-ALM Method Summary n° 170 : DNA methylation analysis using HumanMethylation450k microarray

Mechanistic Studies, Systemic Toxicity, Cancer Research

Genome-wide DNA methylation profiles can be obtained, assessed and comparatively evaluated across different human human cells, tissues, and treatments using Illumina 450K BeadChip Array. The results can be used in the assessment of the xenobiotic effects on epigenome.

1. General Information

Corresponding author(s):

Bioinformatician Pavlo Lutsik, Saarland University

Dr. Sascha Tierling, Saarland University

Prof. Dr. Jörn Walter, Saarland University

Date of data sheet creation : 14 January 2016

Last version number : 1

Last update : 14 January 2016

Proprietary and/or Confidentiality Issues:

The method is based on the Illumina 450K BeadChip kit, which is developed, manufactured and commercially distributed for research use only by Illumina, Inc. All main method components including the BeadChip Array, iScan or HiScan array scanners and GenomeStudio software are registered trademarks of Illumina, Inc.

Assay Throughput:

Maximum 72 samples can be processed in 3 days.

Status

Known Laboratory Use:

The HumanMethylation450k Microarray method is widely accepted by the scientific community because of its representative character and its cost efficiency. It is one of the most popular techniques for genome-wide profiling of DNA methylation, with thousands of references to its various applications in the scientific literature and public repositories (*on 17 Feb 2016 Google Scholar, queried with "Illumina DNA Methylation 450k Array" counted 1,870 results; EBI-EMBL ArrayExpress contained processed data from 464 studies, uploaded since 2011*).

Participation in Evaluation Studies:

The performance of the HumanMethylation450k Microarray was evaluated in comparison with (i) an earlier generation of the HumanMethylation27 Microarray, (ii) reduced representation- and (iii) whole-genome bisulfite sequencing. In all cases a very good correlation was reported (Bibikova et al., 2011; Sandoval et al., 2011; Dedeurwaerder et al., 2011).

The potential of the HumanMethylation450k Microarray for hepatotoxicity profiling was evaluated in the 7th European Framework Programme NOTOX project (notox-sb.eu). HepaRG cells were treated with acetaminophen and valproic acid, DNA was isolated and profiled on 450K BeadChip arrays. Comparative studies including statistics were performed. The method was considered very suitable for such representative comparative studies.

NOTOX

2. Method Definition

Purpose of the Method

TYPE OF TESTING	: Part of an integrated testing strategy
LEVEL OF TOXICITY ASSESSMENT	: Risk assessment
PURPOSE OF TESTING	: Mechanistic studies, Screen or priority setting

Context of use

Currently the method is used for **Non-regulatory** purposes, mainly in basic research: systems biology studies of human cancers and their biomarkers, cell differentiation and pathology (Sandoval et al., 2011).

Within the NOTOX project the method was used to profile primary liver cells and related cell lines (HepaRG, HepG2), as potential candidate to be included in the integrated testing strategy for systemic toxicity. This method summary focuses only on the DNA sample analysis. Neither the cell culture or exposure are considered.

Scientific Principle of the Method

DNA methylation is an essential element of the genome maintenance, global regulation of gene expression. It defines cell identity, development and epigenetic memory (Bird, 2002). Molecular systems which introduce and maintain methylation of specific cytosine residues in the DNA are both dynamic and essential in most organisms. The CpG sequences are targeted most frequently and it is estimated that up to 80% of all CpGs are methylated in an adult mammalian genome (Ziller et al., 2013). The changes in the methylation patterns are observed both during normal development and disease (especially cancer) and can be interpreted as a general marker of a perturbation of a biological system (Roberts, 2005).

Whole genome sequencing of mammalian genomes to identify epigenetic methylation marks is very cost- and labour-intensive. The presented method is based on a multiplexed primer extension approach on approx. 485.000 CpG dinucleotides (~2% of all CpGs in the genome) to catch the information of the most relevant epigenetically regulated regions (promoters, enhancers; mostly obtained from cancer studies). Fluorescence-based signals are detected and compared to obtain a relative methylation beta value for each CpG position providing single-base methylation data.

Comparative analysis between samples can be done using the RnBeads software package developed in the group of Dr. Tierling at Saarland University.

Procedure Description

Biological Endpoint and Measurement:

Biologically, CpG methylation is detected from 485.000 CpGs distributed throughout the genome (predominantly promoters, enhancers).

The measurement is based on fluorescence signals, one for methylated DNA, one for unmethylated DNA. Relative comparison of both signals from approx. 485.000 CpG positions is output as methylation beta value.

Endpoint Value:

BETA VALUE (β) uses the ratio of intensities between methylated and unmethylated alleles to estimate the methylation level of the CpG locus.

Experimental System:

The Infinium HumanMethylation450 BeadChip array containing predefined oligonucleotides bound to

beads. Bisulfite-treated DNA binds to the complementary oligos on the array. Subsequently hybridised oligos are extended in a methylation-dependent manner by one single, fluorophore-marked nucleotide.

Metabolic Competence:

Not present in the experimental setup. Related to the source material from which DNA samples are obtained.

Basic procedure:

Genomic DNA is extracted from the source material (eg. chemical-treated or untreated HepaRG or HepG2 cells). DNA is checked for purity and integrity, quantified and treated with bisulfite. Bisulfite-treated DNA is analysed with the Infinium HumanMethylation450 BeadChip array, according to the manufacturer's instructions (Bibikova et al., 2011). Afterwards, high resolution images of fluorescent-labeled arrays are captured with an array scanner and analysed with dedicated software.

Quality/Acceptance Criteria

The quality control steps and acceptance criteria for a valid experimental run are defined by the developer of Infinium HumanMethylation450 BeadChip array and made available with a purchased kit and dedicated analytical equipment and software. They include checks of the purity and integrity of sample DNA, bisulfite treatment efficiency and internal controls for the performance of the BeadChip array (hybridisation, readout and normalisation steps) (Bibikova et al., 2011; Assenov et al., 2014).

Quality of the results mostly depends on input DNA quality and bisulfite treatment efficiency.

Data Interpretation and/or Prediction Models

Data Analysis:

The assay readout is the methylation beta-value, calculated for each of the 485,577 potential methylation sites, represented on the array as beads carrying probes specific for the methylated and the unmethylated alleles.

$\beta = \text{intensity of the Methylated allele (M)} / [\text{intensity of the Unmethylated allele (U)} + \text{intensity of the Methylated allele (M)} + 100]$

The assay is designed in such way that the oligonucleotide probes for methylated and unmethylated alleles are either located on (i) two different beads and labeled with one fluorescent dye (*Infinium I BeadChip*) or (ii) on one bead and labeled with two different fluorescent dyes (*Infinium II BeadChip*) (Dedeuwaerder et al., 2011). In both cases the intensity of the fluorescence is directly proportional to the amount of sample DNA bound to the probe. Hence the unmethylated CpG's will have $\beta=0$, hemi-methylated $\beta \approx 0.5$, methylated $\beta=1$.

Data analysis process consists of several steps: image processing to identify the location and identity of the individual beads, matching their probe sequences (array internal controls or genomic loci), correction for background, positional and batch effects, signal normalisation, calculation of β value for each allele pair, and finally the differential analysis of DNA methylation patterns across arrays. This can be accomplished with the Methylation Module of the GenomeStudio software, developed and distributed by Illumina Inc.

Alternatively, the RnBeads software package can be used, developed for the open source R software environment for statistical computing (Assenov et al., 2014). Raw data from Infinium 450K experiments can be imported directly into RnBeads to conduct preprocessing, normalization, quality check and differential methylation analysis. In addition the RnBeads can be used to analyse DNA methylation data for different platforms: bisulfate sequencing experiments and HumanMethylation27 Microarray. It enables differential methylation analysis with publicly available DNA methylation datasets, eg. deposited in the Gene Expression Omnibus (GEO) database (Assenov et al., 2014).

Prediction Model (PM):

Not applicable.

3. Method Performance

Robustness of the Method

Within-laboratory reproducibility:

The within-laboratory reproducibility is considered very high. The reported average correlation of β -values between technical replicates ranges from $R^2=0.992$ to $R^2=0.998$ (Bibikova et al., 2011; Dedeurwaerder et al., 2011).

Between-laboratory transferability and reproducibility:

Formal between-laboratory transferability studies, as defined in the ECVAM's modular approach to validation (OECD, 2005), have never been conducted for the use of the Infinium HumanMethylation450 BeadChip array. However, since its introduction in 2011, the method has been adopted in dozens of laboratories worldwide and became a widely accepted standard in the DNA methylation profiling. Standardised array design, high quality manufacturing and extensive data analysis process allow for combining various datasets even if they were obtained in different experiments and by different laboratories.

One of the largest multilaboratory initiatives of its kind is the **The Cancer Genome Atlas (TCGA)**, a coordinated effort to accelerate the understanding of the molecular basis of cancer through the application of genome analysis technologies, including large-scale genome sequencing and epigenetic profiling. The consortium has generated a large collection of cancer methylomes using the Infinium 450k assay, performed in different batches and in different laboratories. The **RnBeads** analysis of 4034 blood, breast, intestinal and brain cancer samples showed data consistency across the dataset and the feasibility of re-assessing combined results from different studies for a new purpose (Assenov et al., 2011).

Test Compounds

Predictive capacity:

Not applicable

Applicability:

Any eventual limitations would be related to the experimental design and specific cell or tissue type tested, which are beyond scope of this summary.

4. Discussion

Known experimental strengths and/or limitations of the method

- **Strengths:**
The method is fast, reproducible and cost-effective (app. 250 €/sample) and easy to perform, when compared to alternatives, such as bisulfite sequencing. Only 250ng of DNA is required and no amplification is needed. Large user base and many data sets from various studies are publicly available for all common cell lines, tissues and many cancers. They can be re-used as reference set, in hypothesis testing or aid in planning of a new experiments.
- **Limitations:**
Initial investment needed in an array scanner, computational resources and training.
The probe coverage of the Infinium HumanMethylation450 BeadChip is not exhaustive. Only approximately 485.000 CpGs are evaluated, out of 28 million estimated possible methylation sites in the genome (Assenov et al., 2011). Some elements, known to be methylated in the human genome (eg. repetitive elements) are not included, so that some important epigenetic information might not be detected.

Technical requirements

- 250 ng of genomic DNA is needed
- BeadChips have to be purchased from Illumina

- iScan or HiScan machine has to be purchased from Illumina

Potential for future development

The method is undergoing continuous development. New EPIC array is already released by Illumina now covering 850.000 CpG positions. It has been shown to be highly reproducible at the 450K CpG sites and consistent among technical replicates while offering additional coverage of 333,265 CpGs located in enhancer regions identified by the ENCODE and FANTOM5 projects (Moran et al., 2016).

Comparisons to other methods

Validation of 450K BeadChip array results using local deep sequencing revealed high confidence of 450K data. It is considered as one of the best established techniques for DNA methylation analysis. For a more general review of other techniques refer to eg. a recently published comparative review by Kurdyukov and Bullock (2016).

5. Contact Details

Bioinformatician Pavlo Lutsik

PhD student
FR8.3 Life Sciences, Genetics/Epigenetics
Saarland University
Function: Corresponding author, Method developer / user
Campus, Bld. A2 4
Saarbrücken D-66123
email: p.lutsik@mx.uni-saarland.de
telephone: +49 681 302 2881

Dr. Sascha Tierling

Post-Doc, Biologist
FR8.3 Life Sciences, Genetics/Epigenetics
Saarland University
Function: Corresponding author, Method developer / user
Campus, Bld. A2 4
Saarbrücken D-66123
email: s.tierling@mx.uni-saarland.de
telephone: Tel: +49 681 302 3295

Prof. Dr. Jörn Walter

FR8.3 Life Sciences, Genetics/Epigenetics
Saarland University
Function: Corresponding author
Campus, Bld. A2 4
Saarbrücken D-66123
email: j.walter@mx.uni-saarland.de
telephone: +49 (0) 681 302 4367

Ver n°	Date	Name	Type of change
1	14 January 2016	Dr. Sascha Tierling	Summary created

6. Abbreviations & Definitions

450K BeadChip Array: Microarray technology developed and distributed by Illumina

BETA VALUE (β): $\beta = \text{intensity of the Methylated allele (M)} / [\text{intensity of the Unmethylated allele (U)} + \text{intensity of the Methylated allele (M)} + 100]$

CpG: cytosine and guanine separated by a phosphate group

DNA: Deoxyribonucleic acid

7. Bibliography

Method development:

- Assenov Y, Müller F, Lutsik P, Walter J, Lengauer T, Bock C. (2014)
Comprehensive analysis of DNA methylation data with RnBeads.
Nature Methods 11(11), 1138-40
- Bibikova, M., Barnes, B., Tsan, C., Ho, V., Klotzle, B., Le, J.M., Delano, D., Zhang, L., Schroth, G.P., Gunderson, K.L., et al. (2011)
High density DNA methylation array with single CpG site resolution.
Genomics 98, 288–295
- Dedeurwaerder, S., Defrance, M., Calonne, E., Denis, H., Sotiriou, C., and Fuks, F. (2011)
Evaluation of the Infinium Methylation 450K technology.
Epigenomics 3, 771–784
- Sandoval J, Heyn H, Moran S, Serra-Musach J, Pujana MA, Bibikova M, Esteller M. (2011)
Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome.
Epigenetics 6(6), 692-702

Last update: 14-01-2016

DB-ALM Protocol n° 170 : DNA methylation analysis using HumanMethylation450 microarray

Mechanistic Studies

A detailed protocol for sample preparation when conducting epigenomic evaluations for all scientists interested in rough epigenetic profiling. The data are obtained, assessed and comparatively evaluated with the untreated controls using Illumina 450K BeadChip Array.

Résumé

Epigenomic data are obtained, assessed and comparatively evaluated with the untreated controls using Illumina 450K BeadChip Array. With this primer extension-based array up to 485.000 CpG positions can be scanned covering promoters and gene body regions. The described method is ideal for rough estimations about DNA methylation changes between two or more samples. DNA methylation of treated and control samples is studied on a genome-wide representative microarray. Epigenomic evaluation is performed on the Illumina 450K BeadChip array. It covers up to 485.000 CpGs spread over the genome (mostly in promoters and gene bodies). The method is cheap (~250 €/sample) and easy to perform. The limitations are the limited coverage of CpGs (repetitive elements are missing).

More information about the method is included in the [DB-ALM Method summary n° 170 : DNA methylation analysis using HumanMethylation450k microarray](#).

Experimental Description

Endpoint and Endpoint Measurement:

Biologically, CpG methylation is detected from 485.000 CpGs distributed throughout the genome (predominantly promoters, enhancers).

The measurement is based on fluorescence signals, one for methylated DNA, one for unmethylated DNA. Relative comparison of both signals from approx.. 485.000 CpG positions is output as methylation beta value.

Endpoint Value:

BETA VALUE (β) uses the ratio of intensities between methylated and unmethylated alleles to estimate the methylation level of the CpG locus.

Experimental System(s):

The Infinium HumanMethylation450 BeadChip array containing predefined oligonucleotides bound to beads. Bisulfite-treated DNA binds to the complementary oligos on the array. Subsequently hybridised oligos are extended in a methylation-dependent manner by one single, fluorophore-marked nucleotide.

Health and Safety Issues

General Precautions

Sodium bisulfite is potentially inducing mutations in DNA – please wear gloves when performing bisulfite treatments with DNA. During the microarray a formamide-containing solution is used, which is potentially toxic and, in case of improper handling, can be harmful. Therefore, all the operations involving formamide should be carried out with all necessary precautions by specially trained staff.

Abbreviations and Definitions

450K BeadChip Array: Microarray technology developed and distributed by Illumina Inc.

BETA VALUE (β): $\beta = \text{intensity of the Methylated allele (M)} / [\text{intensity of the Unmethylated allele (U)} + \text{intensity of the Methylated allele (M)} + 100]$

CpG: cytosine and guanine separated by a phosphate group

DNA: Deoxyribonucleic acid

IDAT: proprietary data format is used to store BeadArray data from the genomewide profiling platforms from Illumina Inc.

R: a programming language and software environment for statistical computing and graphics

VPA: vaploic acid

Last update: 14 January 2016

PROCEDURE DETAILS, 14 January 2016

DNA methylation analysis using HumanMethylation450 microarray DB-ALM Protocol n° 170

The protocol was developed within the NOTOX EU integrated project (*Seurat-1*)

Contact Details

Bioinformatician Pavlo Lutsik

PhD student
FR8.3 Life Sciences, Genetics/Epigenetics
Saarland University
Campus, Bld. A2 4
Saarbrücken D-66123
email: p.lutsik@mx.uni-saarland.de
telephone: +49 681 302 2881

Dr. Sascha Tierling

Post-Doc, Biologist
FR8.3 Life Sciences, Genetics/Epigenetics
Saarland University
Campus, Bld. A2 4
Saarbrücken D-66123
email: s.tierling@mx.uni-saarland.de
telephone: Tel: +49 681 302 3295

Materials and Preparations

Cell or Test System

HepaRG cells, primary hepatocytes or any other cell line or tissue can be used

Equipment

Fixed Equipment

- Illumina iScan or HiScan machines
- NanoDrop 2000c (Thermo Scientific)

Consumables

- GenElute™ Mammalian Genomic DNA Miniprep Kit (Sigma)
- EZ DNA Methylation™ Kit (Zymo Res.)
- Illumina HumanMethylation450 BeadChip Array

Media, Reagents, Sera, others

DMEM with particular supplements (can be obtained from eg. Biopredic Int.)

Preparations

Test Compounds

Valproic acid , acetaminophen

Positive Control(s)

Control oligos on the 450K BeadChip array

Negative Control(s)

Untreated samples

Method

Routine Culture Procedure

Cells were scraped from one well of a 6-well plate, centrifuged for 5 minutes at 900g and washed once

with 1xPBS. Then cells were split equally in two aliquots and frozen at -80°C until DNA/RNA preparation.

Test Material Exposure Procedures

HepaRG cell line exposed to 0.05 and 0.15 mM valproic acid up to 21 days and 1 mM acetaminophen up to 1 day. Primary hepatocytes were exposed to 1 mM acetaminophen for up to 21 days.

Cells were scraped from one well of a 6-well plate, centrifuged for 5 minutes at 900g and washed once with 1xPBS. Then cells were split equally in two aliquots and frozen at -80°C until DNA/RNA preparation.

Endpoint Measurement

1. DNA is prepared from one HepaRG cell aliquot using the GenElute™ Mammalian Genomic DNA Miniprep Kit (Sigma) according to the manufacturer's protocol.
2. DNA is eluted in 200 µl ddH₂O, concentrated by speed vac to a final volume of 20 µl and measured by NanoDrop 2000c (Thermo Scientific).
3. DNA integrity is checked by loading 250 ng on a 1.2% agarose gel.
4. 500 ng of genomic DNA were subjected to bisulfite treatment using the EZ DNA Methylation™ Kit (Zymo Res.) according to the manufacturer's protocol.
5. Bisulfite-treated DNA is analysed on **Infinium HumanMethylation450 BeadChip array in the iScan machine (Illumina)**.
6. Raw data were processed and evaluated using the RnBeads software (Assenov et al., 2014).

Acceptance Criteria

Quality checks include the check of bisulfite conversion and the evaluation of control regions on the BeadChip.

Quality of the results is mostly dependent on input DNA quality and bisulfite treatment efficiency.

Data Analysis

Raw HumanMethylation450 data is obtained from the iScan control software (Illumina) as IDAT files. The data is then loaded into R statistical environment and initially preprocessed using the RnBeads package (Assenov et al., 2014).

The raw intensities are normalized using the manufacturer recommended scaling procedure, implemented as method "illumina" in Bioconductor package methylumi (Davis et al., 2015).

The normalized data is further analyzed using the RnBeads default pipeline to generate an HTML analysis report.

Each called CpG position is tested for association with the VPA treatment and cultivation variable independently for each combination of the VPA concentration and cultivation time, comparing to the corresponding control samples.

Prediction Model

Not Applicable

Annexes

Application note:

http://res.illumina.com/documents/products/appnotes/appnote_dna_methylation_analysis_infinium.pdf

Bibliography

- Assenov Y, Müller F, Lutsik P, Walter J, Lengauer T, Bock C. (2014) Comprehensive analysis of DNA methylation data with RnBeads. *Nature Methods* 11(11), 1138-40
- Davis S, Du P, Bilke S, Triche T, Jr. and Bootwalla M (2015) methylumi: Handle Illumina methylation data. R package version 2.17.0.
- Sandoval J, Heyn H, Moran S, Serra-Musach J, Pujana MA, Bibikova M, Esteller M. (2011) Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome.

Epigenetics 6(6), 692-702

DB-ALM Protocol n° 171 : Analysis of histone modifications via chromatin immunoprecipitation and sequencing (ChIP-seq)

Mechanistic Studies

A detailed protocol for ChIP-seq sample preparation when conducting epigenomic evaluations for all scientists interested in genome-wide histone modification analysis. The protocol is applicable to any histone modification of interest.

Résumé

ChIP-seq technique is the current state of the art regarding histone modification profiling on genome-wide scale.

There is a wide range of commercially available antibodies validated for ChIP-seq allowing a deep characterization of histone modification landscapes. The authors studied genome-wide histone modifications of treated and control samples using chromatin immunoprecipitation followed by deep sequencing (ChIP-seq). ChIP-seq data yield valuable information for the characterization of regulatory genomic regions.

The ChIP-seq protocol described here (buffers and solutions) is based on:

Walter K., Lichtinger M., 2011 Chromatin-Immunopräzipitation (ChIP). In: Jansohn M, Rothhämel S (Eds.), Gentechnische Methoden: eine Sammlung von Arbeitsanleitungen für das molekularbiologische Labor., Spektrum Akademischer

Experimental Description

Endpoint and Endpoint Measurement:

Measurement is based on antibody-enriched sequencing depth, normalizing and comparing read numbers. The comparison is made between treated and untreated cells, eg. HepaRG

Experimental System(s):

Chromatin Immuno-precipitates analysed with high-throughput sequencing on the HiSeq2500 machine (Illumina)

Discussion

ChIP-seq data yield valuable information for the characterization of regulatory genomic regions and it is considered a golden standard for genome-wide histone modification profiling.

One known limitation of ChIP-seq is the bias introduced by sonication-assisted chromatin fragmentation.

Status

Known Laboratory Use:

The method is widely accepted by the scientific community as golden standard for genome-wide histone modification profiling.

Health and Safety Issues

General Precautions

During the fixation step 1% formaldehyde solution is used, which is a highly toxic substance and, in case of improper handling, can be harmful. Therefore, all the operations involving formaldehyde should be carried out with all necessary precautions by specially trained staff.

Abbreviations and Definitions

ChIP-seq: chromatin immunoprecipitation followed by deep sequencing

PIC: Protease Inhibitor Cocktail

Last update: 14.01.2016

PROCEDURE DETAILS, 14 January 2016**Analysis of histone modifications via chromatin immunoprecipitation and sequencing (ChIP-seq)
DB-ALM Protocol n° 171**

The protocol was developed within the *NOTOX* EU integrated project (*Seurat-1*)

Contact Details**Dr. Sascha Tierling**

Post-Doc, Biologist
FR8.3 Life Sciences, Genetics/Epigenetics
Saarland University
Campus, Bld. A2 4
Saarbrücken D-66123
email: s.tierling@mx.uni-saarland.de
telephone: Tel: +49 681 302 3295

Biologist Kathrin Kattler

PhD student
FR8.3 Life Sciences, Genetics/Epigenetics
Saarland University
Campus, Bld. A2 4
Saarbrücken D-66123
email: kathrin.kattler@googlemail.com
telephone: +49 681 302 2443

Materials and Preparations**Cell or Test System**

HepaRG cells, treated with Valproic Acid

Equipment*Fixed Equipment***Instrument**

Bioruptor® NGS System
1.5 ml TPX microtubes
DynaMag™
Agilent 2100 Bioanalyzer
cBot
HiSeq™ 2500
DynaMag™

Supplier

Diagenode
Diagenode
Invitrogen
Agilent Technologies
Illumina
Illumina
Invitrogen

*Consumables***Product**

1.5 ml TPX microtubes
H3K4me3 polyclonal antibody
H3K9ac polyclonal antibody
Qubit® dsDNA HS Assay Kit
Bioanalyzer High Sensitivity DNA Kit
TruSeq ChIP Sample Preparation Kit
TruSeq v2 LT Sample Prep Kit PCR Box
TruSeq PE Cluster Kit v3-cBot-HS
TruSeq SBS Kit v3 - HS (200 Cycles)

Supplier

Diagenode
Diagenode
Diagenode
Invitrogen
Agilent Technologies
Illumina
Illumina
Illumina
Illumina

Media, Reagents, Sera, others

DMEM medium plus supplements (can be obtained from Biopredic Int.)

Preparations

Media and Endpoint Assay Solutions

ChIP A

10 mM HEPES-NaOH
10 mM EDTA
0.5 mM EGTA
0,25 % (v / v) Triton X-100
PIC

ChIP B

10 mM HEPES-NaOH
200 mM NaCl
1 mM EDTA
0.5 mM EGTA
0.01 % (v / v) Triton X-100
PIC

ChIP buffer

25 mM Tris-HCl
150 mM NaCl
2 mM EDTA
1 % (v / v) Triton X-100
0.25 % (w / v) SDS
PIC

ChIP dilution buffer

25 mM Tris-HCl
150 mM NaCl
1 mM EDTA
1 % (v / v) Triton X-100
0.75 % (v / v) Glycerol
PIC

ChIP washing buffer 1

20 mM Tris-HCl
150 mM NaCl
2 mM EDTA
1 % (v / v) Triton X-100
0.25 % (w / v) SDS

ChIP washing buffer 2

20 mM Tris-HCl
500 mM NaCl
2 mM EDTA
1 % (v / v) Triton X-100
0.25 % (w / v) SDS

ChIP washing buffer 3

10 mM Tris-HCl
250 mM LiCl
1 mM EDTA
0.5 % (v/v) NP-40
0.5 % (w/v) Sodiumdesoxycholate
TE / NaCl
10 mM Tris-HCl

50 mM NaCl
1 mM EDTA

ChIP elution buffer (freshly prepared)

100 mM NaHCO₃
1 % (w / v) SDS
NaPO₄ buffer
(93 : 7; v / v)
1 M Na₂HPO₄
1 M NaH₂PO₄
PIC
Complete Protease Inhibitor Cocktail Tablets (Roche)

Test Compounds

Valproic acid

Positive Control(s)

none

Negative Control(s)

Untreated cells; non-antibody enriched ChIP fraction

Method

Routine Culture Procedure

Chromatin Preparation

1. Cells are detached by trypsination from each well of a 6-well plate. For each replicate, two wells were pooled resulting in approximately 5×10^6 cells.
2. After fixation with 1% formaldehyde for 10 minutes, cross-linking is stopped by adding glycine to a final concentration of 125 mM and incubation for 5 minutes at room temperature. Samples are kept on ice or at 4° C.
3. Centrifugation steps throughout the chromatin preparation are performed at 800 x g for 5 minutes at 4° C. After washing twice with 125 mM glycine in PBS and once in 1 x PBS, cells are resuspended consecutively in 5 ml ChIP buffer A, respectively ChIP buffer B, and rotated for 10 minutes.
4. After centrifugation, pellets are resuspended in 500 µl ChIP buffer and kept on ice for 10 minutes. Sonication of chromatin is performed in 166 µl volumes in 0.5 ml shearing tubes using the Bioruptor NGS (Diagenode) at 4° C on high power setting with 30 (+ X) cycles consisting of 30 seconds sonication and 30 seconds recovery time. Samples with appropriate size distribution were diluted with double volume ChIP dilution buffer. 20 µl of diluted samples were stored at – 20° C until later use as *Input controls*.

Immunoprecipitation

1. Protein G beads (Invitrogen) are washed twice with 100 mM NaPO₄ buffer and resuspended in 10 µl 100 mM NaPO₄. 2 µl antibody (polyclonal antibodies against H3K4me3 and H3K9ac (Diagenode)) as well as 5 µl BSA (10 mg / ml) are added.
2. After rotation at 4° C for 2 hours, 300 µl sheared chromatin (equivalent to 1 million cells) are added.
3. Immunoprecipitation is conducted overnight. Afterwards supernatant is completely removed and beads are consecutively washed with 1 ml of ChIP wash buffer 1, ChIP wash buffer 2 (twice), ChIP wash buffer 3 and TE / NaCl buffer. All washing steps are performed under rotation at 4° C for 10 minutes.
4. Samples are transferred into new tubes and incubated with 100 µl freshly prepared elution buffer for 15 minutes at room temperature in a shaking device.
5. Eluates are collected in new tubes. Elution is repeated and both eluates are subsequently combined. Besides, 20 µl of saved *Input controls* were mixed with 180 µl elution buffer and carried along.
6. To each sample 4 µl 5 M NaCl, 2 µl 0.5 M EDTA and 2.5 µl Proteinase K (20 mg / ml) are added. Reversion of cross-linking is carried out in a shaking device at 55° C over night.
7. After adding 2 µl 1 M Tris-HCl, DNA is extracted with 200 µl PCI solution (25 / 24 / 1; v / v / v) under

rotation at room temperature for 10 minutes. Samples are centrifuged at 13000 rpm for 15 minutes and the aqueous phase was transferred into a new tube. Extraction is repeated likewise with 200 µl CI solution (24 / 1; v / v).

8. DNA is precipitated with 20 µl 3 M NaOAc, 1 µl glycogen (20 mg / ml) and 500 µl ice-cold absolute ethanol at – 20° C over night.
9. Samples are centrifuged at 13000 rpm at 4° C for 40 minutes, pellets are washed with 700 µl ice-cold 70 % ethanol. After centrifugation for 20 minutes pellets are dried at room temperature.
10. *Input control* samples are dissolved in 50 µl, Immunoprecipitated DNA in 20 µl ddH₂O.
11. DNA concentration is measured with the Qubit® dsDNA HS Assay Kit (Invitrogen) using 2 µl sample.

DNA sequencing

Up to 10 ng DNA are applied for library preparation using the Truseq ChIP Sample Preparation Kit (Illumina).

1. End Repair, 3'-Adenylation, adapter ligation as well as enrichment PCR (using 8 cycles (replicate 1) or 9 cycles (replicate 2)) are conducted strictly according to the manufacturer's protocol.
2. Size selection is achieved using Ampure XP beads (Beckman Coulter) after adapter ligation (replicate 1) or more effectively before end repair reaction (replicate 2). For selection of DNA fragments between 100 bp and 500 bp, 0.5X beads are added to the sample.
3. After 10 minutes, samples are magnetically separated and supernatant is transferred into a new tube.
4. 1.3X beads are added to the supernatant. After 10 minutes incubation, the supernatant is discarded and beads are washed twice with 80 % ethanol.
5. From the dried beads DNA is eluted using 20 µl resuspension buffer (Illumina).
6. Libraries are sequenced 2 x 100 bp paired-end on the HiSeq2000 (Illumina).

Test Material Exposure Procedures

HepaRG cells were exposed to 0.05mM, 0.15 mM and 0.5 mM valproic acid over 14 and 21 days.

Acceptance Criteria

Fixed samples should not be stored more than 6 month as prolonged storage intensifies sonication biases. Sonication parameters for each cell type should be determined in advance.

Quality checks by performing technical replicates; duplication rate assessment; sequencing of non-antibody-bound fraction to determine background signals.

Data Analysis

Low quality ends (phred score=20) of FastQ format reads were trimmed in addition to adapter removal using Trim Galore (version 0.3.3).

The trimmed reads were aligned to the human reference genome (hs37d5) using GEM mapper (version 1.376 beta).

bamToBed from samtools (version 1.1) was used to convert SAM to BAM format. MarkDuplicate (version 1.115) from picard Tools was used to mark the PCR duplications.

The coverage tracks (bigwig files) were generated by bamCoverage (version 1.5.8.1) from deepTools. MACS (version 2.1.0.20140616) was used for peak calling.

MACS2 together with IDR pipeline were used to assess the reproducibility of the called peaks in case of replicates (d0 and d21).

Bibliography

- O'Geen H, Echipare L, Farnham PJ. (2011)
Using ChIP-seq technology to generate high-resolution profiles of histone modifications. doi: 10.1007/978-1-61779-316-5_20
Methods in Molecular Biology 791, 265-86.
- Yang Y, Fear J, Hu J, Haecker I, Zhou L, Renne R, Bloom D, McIntyre LM (2014)
Leveraging biological replicates to improve analysis in ChIP-seq experiments. *Comput Struct Biotechnol*

J.
9:e201401002.



Physiologically-Based Kinetic models

A Physiologically Based Kinetic (PBK) model is a mathematical model for predicting the absorption, distribution, metabolism and excretion (ADME) of a compound in humans and other animal species. The model can be used to simulate relevant time profiles concentration of selected chemicals and its metabolites using a calibrated PBK model for rat and for human. The present PBK models will be implemented into the KNIME workflow, in order to create an automated tool for Risk Assessment. Further use of the workflow can be in cosmetic Risk Assessment as well for HTS, for in vitro testing design. The KNIME workflow is available via [COSMOS space](#). A web version of the model was also implemented, allowing to execute the model via a web browser. It is available through the [COSMOS KNIME WebPortal](#)..

Method Summary n° 161.....81

The Virtual Cell Based Assay

The Virtual Cell Based Assay (VCBA) is a mathematical model for predicting the fate of a chemical within an in vitro system. The model can be used to simulate relevant intracellular concentrations of selected chemicals in cell lines, and it can be coupled with physiologically based kinetic (PBK) models to predict human in vivo toxic doses from concentrations that cause effects in vitro. The characterization of the concentration that produces an effect (whether this is a perturbation of a molecular pathway or an apical toxic endpoint) is necessary at two levels: first, in analysing the results of vitro experiments, since “nominal” concentrations do not represent the real concentration experienced by the cell; and, second, in extrapolating to humans, since the true concentration experienced by cells within the target organ is more representative for human toxicity assessment. In order to analyse the results of *in vitro* experiments, a VCBA was developed for liver cell lines (3T3cBalbc, HepG2 and HepaRG). The VCBA model has been further extended to include cell lines from different organs: lung A459 cells and from heart cardiomyocytes. The KNIME workflow is available via [COSMOS space](#). A web version of the model was also implemented, allowing to execute the model via a web browser. It is available through the [COSMOS KNIME WebPortal](#).

Method Summary n° 162.....103

The Human Bio-Accumulation Model

The Human Bio-Accumulation Model is a mathematical model based on *in vitro* liver metabolism data, minimal renal excretion and a chronic exposure, to assess the bioaccumulative potential of a chemical. The approach has been analysed using literature data for some well-known bioaccumulative compounds, data from the EURL ECVAM database, and for a subset of the ToxCast phase I chemical library. The aim of the present method was to develop an open source approach for assessing the potential of chemicals (personal care products and cosmetics) to bioaccumulate in humans following realistic exposure scenarios. An important characteristic, addressed in the work of Tonnelier et al. (2012) is to assess chemicals for human bioaccumulative potential using a simple physiologically based toxicokinetic model, which ranks chemicals based on their human bio concentration factor (hBCF). The work on the model was carried out at the Joint Research Centre under the [COSMOS](#) project, SEURAT 1 Research Initiative. The KNIME workflow is available via [COSMOS space](#). A web version of the model was also implemented, allowing to execute the model via a web browser. It is available through the [COSMOS KNIME WebPortal](#).

Method Summary n° 163.....119

PPAR γ full agonists' pharmacophore-based predictor

The PPAR γ full agonists' pharmacophore-based predictor is an *in silico* model to predict the potential of chemicals to bind and activate PPAR γ (PPAR γ full agonism prediction). Taking into consideration that ligand-dependent activation of PPAR γ can be a molecular initiating event (MIE) that triggers downstream events leading to liver steatosis, the model can be used for *in silico* prediction of potential prosteatotic compounds that are PPAR γ full agonists. The implemented model is based on a dataset of 170 PPAR γ ligands extracted from Protein Data Bank (PDB) (www.rcsb.org) and scientific literature. The model is based on a two-step procedure: docking of the chemical structures into the PPAR γ binding site and filtering using a pharmacophore model of PPAR γ full agonists. The work on the model was carried out under the [COSMOS](#) project, SEURAT 1 Research Initiative.

Method Summary n° 168.....131

QSAR model for the prediction of LXR binding potential

This quantitative structure-activity relationship ((Q)SAR) model is a classification model (based on PLS-DA) for the prediction of Liver X Receptor (LXR) binding potential. The binding and activation of LXR has been identified among the molecular initiating events leading to liver steatosis. Thus, this QSAR model may be used as part of integrated approaches to testing and assessment (IATA)/integrated testing strategies (ITS) to screen chemicals potentially sharing the same AOP (i.e. from LXR activation to liver steatosis). The model may also find application in the grouping of chemicals to support read-across justification, in terms of mechanistic plausibility (as molecular initiating event leading to long-term adverse effects in the liver), and weight of evidence evaluations. The KNIME workflow is available via [COSMOS space](#). A web version of the model was also implemented, allowing to execute the model via a web browser. It is available through the [COSMOS KNIME WebPortal](#).

Method Summary n° 169.....137

PPAR γ full agonists' potency predictor

PPAR γ full agonists' potency predictor is a three dimensional QSAR (Quantitative Structure-Activity Relationships) model for prediction of chemical compounds' potency (transactivation activity, pEC₅₀) to induce PPAR γ -mediated expression of target reporter gene. PPAR γ activation by full agonists has been outlined as one of the possible molecular initiating events for mode of action leading to non-alcoholic fatty liver disease (NAFLD) (Al Sharif et al., 2014). The experimental data reported in the literature used for the selected endpoint has been obtained from Luciferase reporter gene assay. The last is relevant to the OECD conceptual framework, presenting the most promising assays to detect and characterise chemical effects on the PPAR signaling pathway and is among the in vitro assays that are going to be considered for incorporation into new or existing Test Guidelines for the detection of endocrine disrupting chemicals after refinement and validation (OECD, ENV/JM/MONO(2012)23). The 3D QSAR model is developed using a data set of 122 full PPAR γ agonists (83 – training set, 39 – test set), selected from a previously constructed and publically available PPAR γ ligands' dataset (<http://biomed.bas.bg/qsarmm/>) with structural and biological (binding affinity, IC₅₀; transactivation activity, EC₅₀; relative efficacy, %max) data for 439 agonists (full and partial). The model predicts transactivation activity using calculated electrostatic, acceptor and hydrophobic similarity indices as 3D descriptors, generated by CoMSIA (Comparative Molecular Similarity Indices Analysis) 3D QSAR approach. The work on the model was carried out under the [COSMOS](#) project, SEURAT 1 Research Initiative.

Method Summary n° 175.....145

Parallel Artificial Membrane Permeability Assay (PAMPA) logP_m predictor

PAMPA logP_m predictor is a mathematical model for prediction of chemical compounds' permeability in Parallel Artificial Membrane Permeability Assay (PAMPA), which is assumed to be a practical estimator of passive gastro-intestinal absorption (GIA) of low molecular weight compounds. The implemented model is developed using a data set of 276 compounds and predicts permeability coefficient P_m using *in silico* calculated distribution coefficient (logD) and the ratio of the Topological Polar Surface Area (TPSA) to Molecular Weight (MW) as descriptors. With a view to the free and open implementation of the models and lack of such tools for logD estimation, two implementations of the model were produced based on logD estimations readily obtainable through free online services. The KNIME workflow is available via [COSMOS space](#). A web version of the model was also implemented, allowing to execute the model via a web browser. It is available through the [COSMOS KNIME WebPortal](#).

Method Summary n° 176.....151

In silico screening for potential nuclear receptor ligands and alerts towards hepatosteatosi

Many pathways in toxicology can be traced back to nuclear receptors (NR) as initial target. One notable example is endocrine disruption, another the development of fatty liver (hepatosteatosi) associated with chronic exposure to the agonists of some nuclear receptors, e.g. Retinoic Acid Receptor, Retinoid X receptor, Liver X Receptor and Peroxisome Proliferator-Activated Receptor (Moya et al 2010). For the purpose of this model the structural and physico-chemical features of NR ligands have been investigated using data from ChEMBL and the Protein Data Bank (PDB). The information obtained was implemented into a KNIME workflow (see Steinmetz et al 2015) and can be used for screening for potential NR ligands during drug development and risk assessment, which can support hazard identification, category formation for read-across and provide knowledge to anchor adverse outcome pathways (AOPs). The KNIME workflow is available via [COSMOS space](#). A web version of the model was also implemented, allowing to execute the model via a web browser. It is available through the [COSMOS KNIME WebPortal](#).

Method Summary n° 177.....157

***In silico* profiler for covalent DNA binding**

One of the fundamental and initial steps for genotoxicity is the formation of a covalent bond between nucleophile and electrophile molecules of DNA and xenobiotic. DNA binding can thus be a molecular initiating event in an adverse outcome pathway (AOP) leading to both mutagenicity and carcinogenicity. The *in silico* profiler for covalent DNA binding was developed based on the review of current scientific knowledge relating to mutagenicity and genotoxic carcinogenicity structural alerts and mapping the existing structural alerts in terms of their relationships with mechanistic organic chemistry, i.e. identifying alerts from the published compilations related to covalent DNA binding. The mapping was performed to achieve maximum overlap and usability whilst restricting redundancy in the alerts, and to ensure that the alerts related to the molecular initiating event of covalent DNA binding. A total of 111 structural alerts, supported by mechanistic information and meta data, have been created. The KNIME workflow is available via [COSMOS space](#). A web version of the model was also implemented, allowing to execute the model via a web browser. It is available through the [COSMOS KNIME WebPortal](#).

Method Summary n° 178.....165

***In silico* profiler for hepatotoxicity**

The *in silico* profiler for hepatotoxicity consists of 16 structural alerts coded into SMARTS patterns. The alerts can be used in the screening of compounds to highlight potential hepatotoxicity. The profiler is envisaged to be used to develop chemical categories based upon similar mechanisms as part of the adverse outcome pathway paradigm. The profiler has also the potential to act as a knowledge generator serving as a starting platform from which to conduct additional toxicological studies. The KNIME workflow is available via [COSMOS space](#). A web version of the model was also implemented, allowing to execute the model via a web browser. It is available through the [COSMOS KNIME WebPortal](#).

Method Summary n° 179.....177

***In silico* profiler for mitochondrial toxicity**

The main goal of the *in silico* profiler is grouping chemicals into mechanism based categories centred on the knowledge of a number of key molecular initiating events (MIEs) for mitochondrial toxicity as part of the adverse outcome pathway (AOP) paradigm. The *in silico* profiler for mitochondrial toxicity consists of 21 structural alerts coded into SMARTS patterns. The KNIME workflow is available via [COSMOS space](#). A web version of the model was also implemented, allowing to execute the model via a web browser. It is available through the [COSMOS KNIME WebPortal](#).

Method Summary n° 180.....187

***In silico* profiler for covalent protein binding**

For a number of toxicological endpoints, such as hepatotoxicity (liver fibrosis), skin sensitisation, respiratory sensitisation or aquatic toxicity, one of the fundamental steps is the formation of a covalent adduct between a reactive exogenous chemical (electrophile) and biological protein (nucleophile). Protein binding can thus be a molecular initiating event for a number of adverse outcome pathways. The *in silico* profiler for covalent protein binding was developed based on the review of current scientific knowledge on structural alerts relating to a number of toxicity endpoints. The KNIME workflow is available via [COSMOS space](#). A web version of the model was also implemented, allowing to execute the model via a web browser. It is available through the [COSMOS KNIME WebPortal](#).

Method Summary n° 181.....199

***In silico* skin permeability estimation**

Dermal absorption is of great importance for safety evaluation of chemicals taken up via dermal route of exposure regarding the availability of the substance in the organism. The purpose of this *in silico* workflow is to estimate the skin permeability coefficient (kp) for an organic compound, based on calculated physico-chemical descriptors. The KNIME workflow is available via [COSMOS space](#). A web version of the model was also implemented, allowing to execute the model via a web browser. It is available through the [COSMOS KNIME WebPortal](#).

Method Summary n° 182.....211

DB-ALM Method summary n° 161 : Physiologically-Based Kinetic models

Systemic Toxicity, Biokinetics

A **Physiologically Based Kinetic (PBK)** model is a mathematical model for predicting the absorption, distribution, metabolism and excretion (ADME) of a compound in humans and other animal species. The model can be used to simulate relevant time profiles concentration of selected chemicals and its metabolites using a calibrated PBK model for rat and for human. The models were programmed in R (free software environment for statistical computing) for further incorporation into KNIME workflows.

1. General Information

Corresponding author(s):

Dr. Andrew Worth, European Commission - Joint Research Centre

Date of data sheet creation: 15 January 2015

Last version number: 1

Last update: 15 January 2015

Section **No.5 Contact Details** provides a complete list of the authors and method developers.

The method summary presented here describes a mathematical model.

In the context of this document, the term "method" in the paragraph headings refers to the PBK model.

Proprietary and/or Confidentiality Issues:

None. Work carried out at the Joint Research Centre with partial support from the FP7 COSMOS project.

Status

Development completed

The PBK models have been built in R and have been translated in KNIME (the Konstanz Information Miner) automated workflows, as a part of the COSMOS (Integrated *In Silico* Models for the Prediction of Human Repeated Dose Toxicity of Cosmetics to Optimise Safety) project, within the SEURAT-1 Research Initiative (Safety Evaluation Ultimately Replacing Animal Testing 1) cluster (2011-2015). More information on the COSMOS project is available on its dedicated website:

<http://www.cosmostox.eu>

The KNIME workflows and documentation are available via COSMOS Space:

<http://cosmosspace.cosmostox.eu>

A web version of the models was also implemented, allowing to execute the models via a web browser. It is available through the COSMOS KNIME WebPortal:

<http://knimewebportal.cosmostox.eu>

A tutorial is available at:

https://www.youtube.com/watch?v=Ihn4l_g981Q/



2. Method Definition

Purpose of the Method

TYPE OF TESTING	:	Part of an integrated testing strategy
LEVEL OF TOXICITY ASSESSMENT	:	Risk assessment
PURPOSE OF TESTING	:	Safety

Context of use

The main goal of this method is to develop and optimise computer-assisted physiologically-based toxicokinetic (PBK) and dynamic (PBD) models to predict the biological and toxicological effects of chemicals in humans under single and repeated exposure conditions. Further use of the workflow can be in cosmetic Risk Assessment as well for HTS *in vitro* experiments.

The modelling approach was illustrated through its application (to selected chemicals found in consumer products and food, as well as some drugs. With a view to the implementation of the models into open-source tools, the models were programmed in **R** for further incorporation into KNIME workflows. See **the R project for statistical computing** for technical documentation:

- <http://www.r-project.org/>

KNIME is a user-friendly graphical workbench for the analysis process such as data access, data transformation, initial investigation, visualization and reporting. It integrates various components for machine learning and data mining through its modular data pipelining concept. Further use of the workflow can be in cosmetic risk assessment as well for high throughput screening (HTS) to facilitate the *in vitro* test design.

- <http://www.knime.org/>

The purpose of the model is to simulate relevant time profiles concentration of selected chemicals and its metabolites using a calibrated PBK model for rat and for human. The present PBK model is implemented into the KNIME workflow, in order to create an automated tool for Risk Assessment. The PBK model (and the this specific KNIME workflow) was adapted to estimate subject-specific concentration-time profiles in human biological samples (serum and urine) with the ultimate goal to back-extrapolate to the time of intake, assuming that the dose ingested is known.

Chemicals used

Coumarin, Estragole, Hydroquinone, Caffeine, Ethanol, Isopropanol, Styrene, Methyl iodide, Nicotine

Target Users and Uses

For in vitro toxicologists:

External dose → concentration in blood or selected organ → → → In vitro testing strategy planning

For Risk Assessors:

External dose → concentration in blood or selected organ → → → Point of Departure (PoD) for extrapolation (Risk Assessment).

Potential Applications

- Predict from *in vitro* to *in vivo*
- Extrapolate from low dose to high dose response
- Extrapolate from route to route exposure
- Build for characterisation of cosmetic ingredients, however it can be applied for consumer products and food components and additives (e.g. coumarin, ethanol and estragole)

Scientific Principle of the Method

Toxicokinetic modelling is currently seen as the most adequate approach to simulate the fate of chemical compounds in the human body (including the possible emergence of metabolites) and their absorption, distribution, metabolism and excretion (ADME) (Adler et al., 2011).

The present PBK model was developed at the JRC to analyse the effects of human exposure to selected chemicals. Physiologically-based toxicokinetic (PBTK) and toxicodynamic (PBDT) models for the healthy adult Caucasian population were constructed and parameterised for the nine case study compounds, which include industrial chemicals and substances found in consumer products and food: coumarin, estragole, hydroquinone, caffeine, ethanol, isopropanol, methyl iodide, styrene and nicotine (Gajewska et al., 2014; 2015; unpublished data).

Model components

The schematic representation of the human PBK model is given in **Figure 1**. A PBK model tries to mimic the human body compartment including the blood circulation to and from the organs and the movement of a chemical substance throughout the compartments and its metabolism as realistically as needed for the purpose of the exercise and as realistically as possible based on constraints with respect to previous knowledge on parameters (in order to prevent what is called an 'overparameterisation').

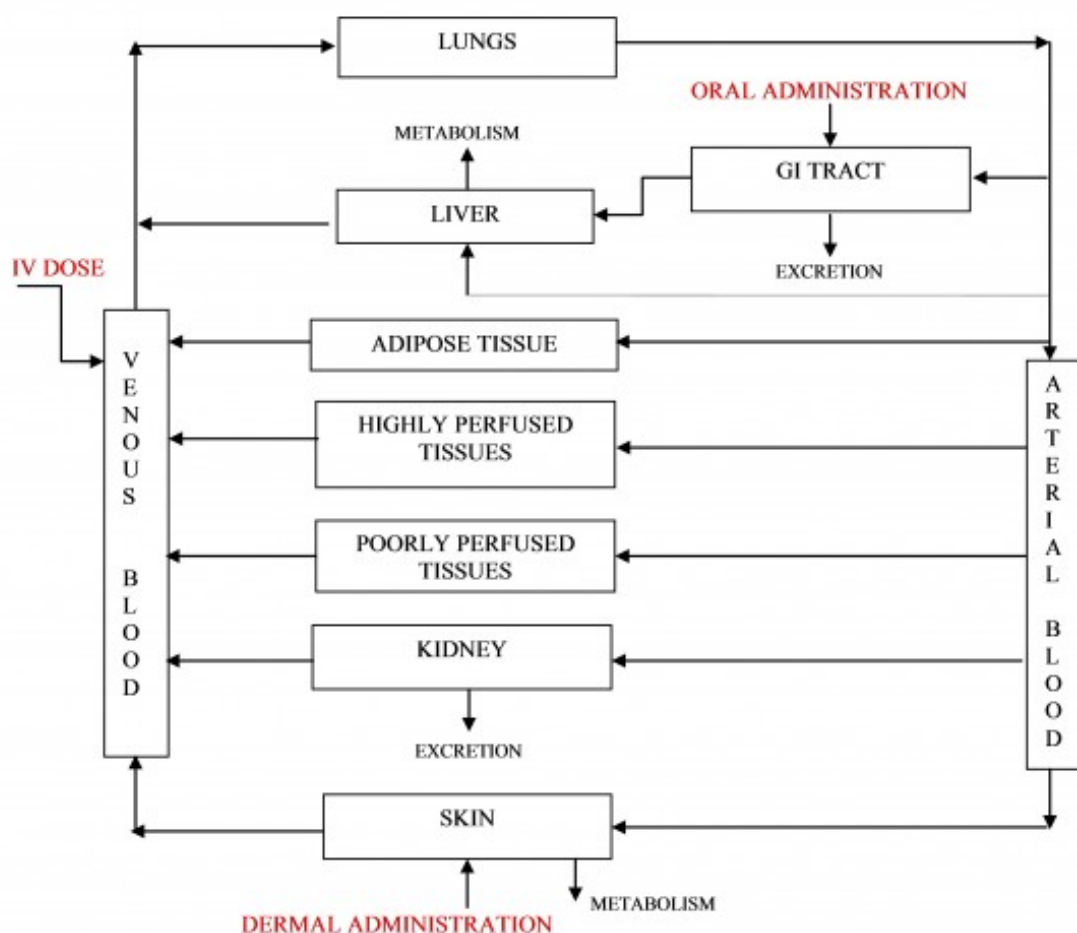


Figure1: General Structure of the PBTK model

Model exposure routes

Exposure Routes used in the present PBK model are: Oral, Dermal and Inhalation (**Figures 2, 3 and 4** below).

In the human PBTK model for oral absorption, the human GI tract is represented by 6 sub-compartments (**Figure 2**). In modelling absorption along the GI tract following gavage administration dissolution from matrix, stomach emptying rate and a first order rate of absorption from stomach, small and large intestine were taken into account.

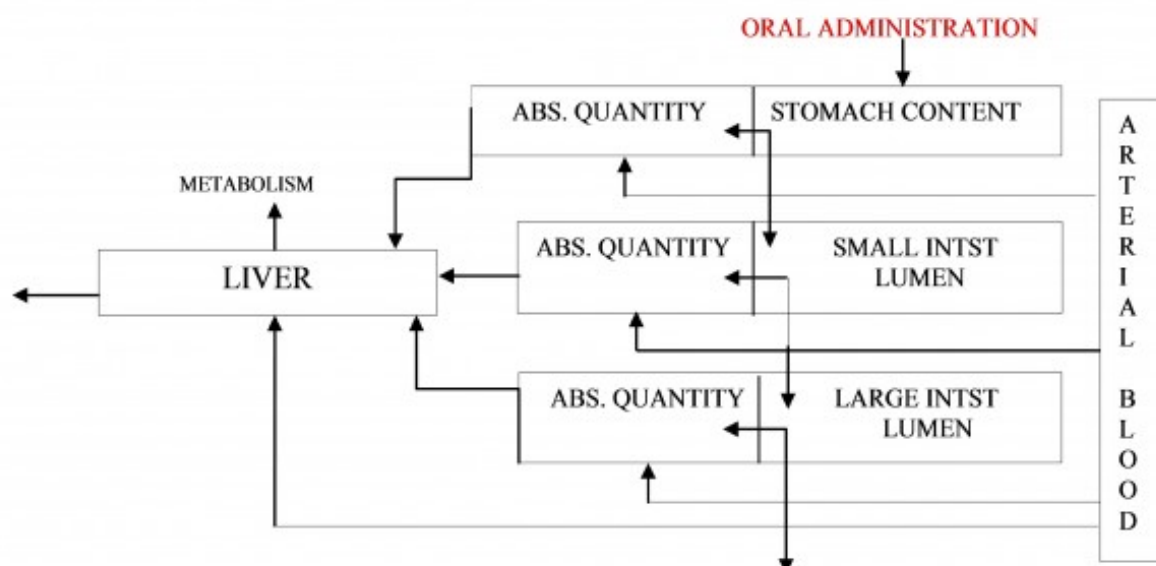


Figure 2: GI tract sub-compartments in human PBK model for oral exposure.

For PBK modelling of human dermal exposure, 4 compartments were used for dermal exposure (surface compartment for the product formulation and 3 skin sub-compartments (stratum corneum, viable epidermis and dermis perfused by blood)). For dermal caffeine exposure, one extra skin sub-compartment had to be added, i.e. hair follicles (see [Figure 3](#)).

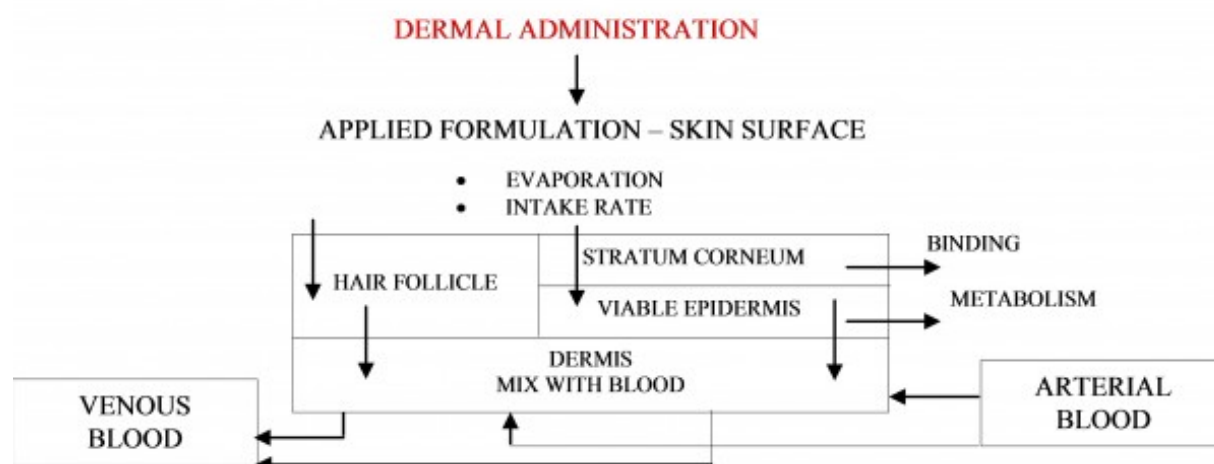


Figure 3: Skin divided into sub-compartments in human PBK model for dermal exposure

The respiratory tract consists of 3 main compartments (24 Weibel generations): convection and diffusion segments (with metabolism that may occur at any stage) ([Figure 4](#)). This dynamic lungs model is more suitable for reactive substances, drugs and nanoparticles. For simple compounds this refinement is not necessary but it allows for a better control of the inhalation/exhalation process. A sinusoidal breathing pattern is incorporated in the model with time frame in seconds (higher computational burden than other exposure types). The model, additionally, consists of blood/air partition coefficient, Ostwald solubility in lungs and rate of dissolved gas removed from lungs due to blood perfusion on overall model prediction performance.

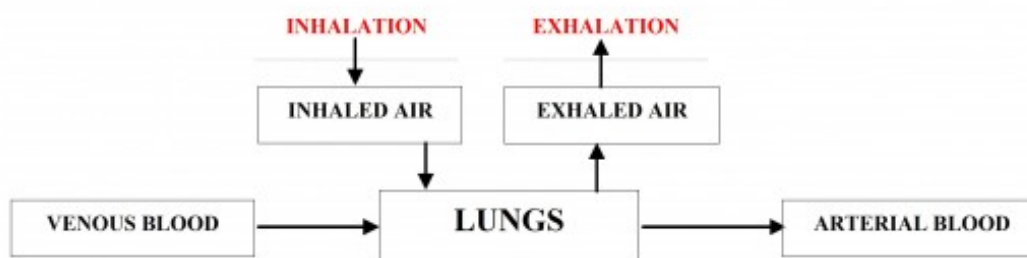


Figure 4: Inhalation in human PBK model

Procedure Description

Biological Endpoint and Measurement:

Physiologically based (pharmacokinetic) model (PBK)

Dose-response

Time-response

Bioavailability

Basic procedure:

The PBK model is a mathematical approach. Examples of the equations used in the individual model components are described in the next section: **2.1 Model Description**. The models for coumarin and caffeine are already coded and implemented in KNIME and can be accessed via COSMOS space:

- <http://cosmospace.cosmostox.eu/app/home>

A presentation on the COSMOS Automated *in silico* tools for *in vitro* to *in vivo* extrapolation was recorded for the American Society for Cellular and Computational Toxicology (ASCCT). Method developer Dr. Alicia Paini explains the composition of the model and the use of the KNIME workflows:

- <https://youtu.be/vJc1UEbuWdE>

2.1 Model Description

Model Equations

The mathematical equations were programmed in R language by combining functionalities of the following R packages available from “The Comprehensive R Archive Network” website (<http://cran.r-project.org>): *deSolve*, *ReacTran*, *PK*, *FME*, *rgenoud* and *AICcmoavg*. Ordinary differential equations (ODEs) were solved by the method *lsoda* available in the *deSolve* package. The method of lines was used to solve partial differential equations (PDEs). Further details are given below.

Model Equation Examples

PBK models were built for several compounds taking into account mode of action of the chemical. We report as an example of mathematical equation describing the processes, the simplest first order rate of absorption **PBK 1** model. However, several PBK models were built describing the absorption, oral and dermal with different sub compartment, for instance:

1. **PBK 3:** GI tract: sub-compartments: stomach content and tissue, small and large intestine content and absorbed quantity
2. **PBK 2:** Skin: 2 compartments: one with Fickian diffusion and one with blood mix.

For further information on these two models (PBK 3, GI tract; and PBK 2, Skin) please consult Gajewska et al. (2014 and 2015).

PBK 1:

GI tract: 1 compartment with first order rate of absorption

$$\text{If } (t > t_{\text{lag}}) \frac{dA_{\text{abs,GI}}}{dt} = -ka_{\text{abs,GI}} \cdot A_{\text{abs,GI}} - kel_{\text{abs,GI}} \cdot A_{\text{abs,GI}}; \text{ else } \frac{dA_{\text{abs,GI}}}{dt} = 0 \quad (\text{A1.1})$$

Skin: 1 compartment with first order rate of absorption

$$\text{If } (t > t_{\text{lag}}) \frac{dA_{\text{abs,skn}}}{dt} = -ka_{\text{abs,skn}} \cdot C_{\text{abs,skn}}; \text{ else } \frac{dA_{\text{abs,skn}}}{dt} = 0; C_{\text{abs,skn}} = \frac{A_{\text{abs,skn}}}{V_{\text{abs,skn}}}; V_{\text{abs,skn}} = \text{Area} \cdot L \quad (\text{A1.2})$$

The rest of compartments/tissues are common for PBK 1, PBK2 and PBK 3:

Adipose tissue (adp), highly perfused tissues (hpt) and poorly perfused tissues (ppt):

$$\frac{dA_{\text{on}}}{dt} = f_{\text{on}} \cdot (C_{\text{art}} - \frac{C_{\text{on}}}{PC_{\text{on}}}) \quad A_{\text{on}}(t=0)=0 \quad C_{\text{art}} = \frac{A_{\text{art}}}{V_{\text{art}}} \quad C_{\text{on}} = \frac{A_{\text{on}}}{V_{\text{on}}} \quad (\text{A1.3})$$

where: on = organ name (adp, hpt, ppt)

Kidney:

$$\frac{dA_{\text{kid}}}{dt} = f_{\text{kid}} \cdot (C_{\text{art}} - \frac{C_{\text{kid}}}{PC_{\text{kid}}}) - \text{CLR} \cdot \frac{C_{\text{kid}}}{PC_{\text{kid}}} \quad (\text{A1.4})$$

Liver (without zonation):

$$\frac{dA_{\text{liv}}}{dt} = \text{flowGItract} + \text{Form}_{\text{liv}} + f_{\text{liv}} \cdot (C_{\text{art}} - \frac{C_{\text{liv}}}{PC_{\text{liv}}}) - \text{MET}_{\text{liv}} \quad A_{\text{liv}}(t=0)=0 \quad C_{\text{liv}} = \frac{A_{\text{liv}}}{V_{\text{liv}}} \quad (\text{A1.5})$$

For a parent compound:

GI tract with 1 compartment:

$$\text{flowGItract} = \frac{dA_{\text{abs,GI}}}{dt}$$

GI tract with sub-compartments:

$$\text{flowGItract} = \text{fra} \cdot f_{\text{git}} \cdot \frac{C_{\text{stm}}}{PC_{\text{git}}} + \text{frb} \cdot f_{\text{git}} \cdot \frac{C_{\text{sl}}}{PC_{\text{git}}} + \text{frc} \cdot f_{\text{git}} \cdot \frac{C_{\text{LI}}}{PC_{\text{git}}}$$

$\text{FORM}_{\text{liv}} = 0$ (rate of formation of metabolites)

MET_{liv} includes all the metabolism equations

Venous blood:

$$\frac{dA_{ven}}{dt} = f_{liv} \cdot \frac{C_{liv}}{PC_{liv}} + f_{ppt} \cdot \frac{C_{ppt}}{PC_{ppt}} + f_{hpt} \cdot \frac{C_{hpt}}{PC_{hpt}} + f_{adp} \cdot \frac{C_{adp}}{PC_{adp}} + f_{kid} \cdot \frac{C_{kid}}{PC_{kid}} + FSKIN - F_{-crd} \cdot C_{ven}$$

$$A_{ven}(t=0)=0 \quad C_{ven} = \frac{A_{ven}}{V_{ven}} \quad (A1.6)$$

For dermal absorption only:

Skin with 1 compartment:

$$FSKIN = \frac{dA_{Abs,skn}}{dt}$$

Skin with 2 compartments:

$$FSKIN = F_{skin} \cdot \frac{C_{skn}}{PC_{skn}}$$

Skin with sub-compartments:

$$FSKIN = \frac{3}{4} \cdot f_{skin} \cdot \frac{C_{skn}}{PC_{skn}} + \frac{1}{4} \cdot f_{skin} \cdot \frac{C_{sknhf}}{PC_{sknhf}}$$

Plasma quantification:

$$C_{ven,PL} = \frac{A_{ven}/(0.55 \cdot V_{ven})}{RBP}$$

Arterial blood and lungs:

$$\frac{dA_{lng}}{dt} = f_{lng} \cdot (C_{ven} - \frac{C_{lng}}{PC_{lng}}) \quad A_{lng}(t=0)=0 \quad C_{lng} = \frac{A_{lng}}{V_{lng}} \quad (A1.7)$$

$$\frac{dA_{art}}{dt} = f_{-crd} \cdot (\frac{C_{lng}}{PC_{lng}} - C_{art}) \quad A_{lng}(t=0)=0 \quad (A1.8)$$

Table 1: Equations used to calculate toxicokinetic parameters

Parameter	Area under curve (AUC)	Bioavailability	Systemic clearance	Volume of distribution (VD)	Elimination (half-life)
Equation	$\int_{t=0}^{t=T_{simul.}} C_{org} \cdot dt$ <p>Where: C_{org} = organ/tissue concentration</p>	$f = \frac{AUC_{ven,t}}{AUC_{ven,IV}}$ <p>Where: ven=venous blood, t = oral, dermal, inhalation; IV= intravenous</p>	$CL = \frac{f \cdot dose}{AUC_{ven}}$	$VD = \frac{\sum_{i=1}^{i=I} AUC_{or,i}}{RBP \cdot AUC_{ven}}$ <p>Where: I = no of organs in the model</p>	$t_{\frac{1}{2}} = \frac{\ln 2 \cdot VD}{CL}$

Input data: Model Parameters

For the human oral model, the most sensitive parameters (GIT dissolution rates, first order uptake rate constants and metabolism parameters for liver) were optimized using measured human data. Liver metabolism parameters were taken from the literature and were optimized with respect to in vivo blood concentrations using human data for the main metabolites only. All physiological parameters for a reference man, woman and rat that are independent of the chemical and constitute a constant part of the model equations are given in Table 2 below. The values were taken from the study of Brown and colleagues (Brown et al., 1997).

Table 2: Physiological system data for human (reference woman and man).

Quantity	Reference woman (standard deviation)	Reference man (standard deviation)
Average body weight [kg]	65 (15)	78 (15)
Organ	Organ weights fractions (fractions of body weight)	
Liver	0.026 (0.006)	0.026 (0.006)
Adipose tissue	0.327 (0.127)	0.213 (0.096)
Lungs	0.0105 (0.002)	0.012 (0.002)
Kidney	0.0044 (0.0004)	0.0044 (0.0004)
GI tract:	0.0265 (0.0095)	0.025 (0.0095)
Stomach:	0.00337	0.00318
Small Intestine:	0.0146	0.0138
Large Intestine:	0.0085	0.0080
Poorly perfused tissues +Skin	0.473(0.07)	0.5855 (0.07)
Highly perfused tissues	0.0676 (0.0025)	0.0621 (0.0025)
Blood	0.065 (0.0035)	0.072 (0.0035)
Venous blood	0.04875	0.054
Arterial blood	0.01625	0.018
L [cm]	0.204 (0.1)	0.2906 (0.1)
L _{ve} [cm]	0.0032 (0.001)	0.0047 (0.001)
L _{sc} [cm]	0.0018 (0.0002)	0.0017 (0.0002)
	Regional blood flow rates (fraction of cardiac output [L/h])	
Cardiac output [L/h]	15 · BW ^{0.74}	
Liver	0.25 (0.05)	0.24 (0.05)
Adipose tissue	0.055 (0.005)	0.04 (0.005)
Lungs	0.025 (0.0009)	0.025 (0.0009)
Kidney	0.19 (0.02)	0.2 (0.02)
Poorly perfused tissues	0.135 (0.03)	0.16 (0.03)
GI tract	0.14 (0.01)	0.13 (0.01)
Fractions of GI tract flow rate that go to Stomach tissue, small Intestine wall and large intestine	0.2, 0.6, 0.2	0.2, 0.6, 0.2
Skin	0.05 (0.01)	0.05 (0.01)
Highly perfused tissues	0.155 (0.015)	0.155 (0.015)

The physicochemical properties for each individual chemical, with parameters to represent the Absorption, Distribution, Metabolism and Excretion (ADME), including partition coefficients and absorption rates, are reported in Gajewska et al. (2014 and 2015).

Model Settings:

- Species : Human
- Dosage: Variable to be set from user (in mg or mg/kg)
- Exposure route: Oral, Dermal, inhalation
- Exposure Time: Variable to be set from user
- Exposure Type: Single or repeat
- Chemical Application type: Variable to be set from user (example, liquid, solid, gel tablet, pill, vehicle for dermal application)

Additional information

A total of 9 PBK models have been developed to date in COSMOS project. The list of available models with respective dosage settings can be found in the summary table below. The models for coumarin and caffeine are implemented in KNIME and can be accessed via COSMOS space.

Table 3: Summary of available models with respective dosage settings.

Compound name	PBTK: oral	PBTK: dermal	PBTK: inhalation	PBTK-PBTD	Reference
Coumarin	single	single	-	single-	Gajewska et al., 2014
Estragole	single	-	-	single-	Unpublished
Hydroquinone	single	single	-	-	Gajewska et al., 2014
Caffeine	single repeated	single	-	single repeated	Gajewska et al., 2014 and 2015
Ethanol	single	single	single	-	Unpublished
Isopropanol	single	single	single	-	Unpublished
Styrene	-	-	single	-	Unpublished
Methyl iodide	-	single		-	Unpublished
Nicotine	single repeated	single repeated	single repeated	single repeated	Gajewska et al., 2014 bis

3. Method Performance

Robustness of the Method

General performance measures:

The PBTK model underwent an extensive evaluation, described in the next section: **3.1 Evaluation of model performance.**

3.1 Evaluation of Model Performance

Sensitivity analysis

Sensitivity analysis and parameter identifiability were performed to identify the most important and sensitive parameters with respect to the blood/plasma AUC for the case study chemicals according to

Soetaert and Petzoldt (Soetaert and Petzoldt, 2010; Soetaert, 2010) prior to their optimization. In local sensitivity analysis (Eq. (1)), all parameters were evaluated individually in a very small region close to their nominal value. A parameter value divided by the average of simulated outputs was used as a scaling factor (SF).

$$\dot{x} = f(x, u, \theta) \quad y = g(x, \theta) \quad S_{ij} = \left(\frac{\theta_{j0}}{SF} \right) \frac{\partial y_i}{\partial \theta_j} \Big|_{\theta=\theta_0} \quad \text{Eq. (1)}$$

$$L1 = \frac{\sum |S_{ij}|}{N} \quad L2 = \sqrt{\frac{\sum (S_{ij}^2)}{N}} \quad \text{Eq. (2)}$$

$$\gamma = \frac{1}{\sqrt{\min(EV \begin{bmatrix} T & \hat{S} \\ S & S \end{bmatrix})}} \quad \text{Eq. (3)}$$

where:

y is vector of function outputs for a specific variable;

x is vector of state variables;

θ is vector of parameters (0 parameter estimate);

u is vector of inputs;

N is number of time points;

\hat{S} is the columns of the sensitivity matrix that correspond to the parameters included in the set;

EV is estimation of the eigenvalues

The following kinetic and compound-specific parameters were analyzed: GI tract absorption rates (D_{ss} (or D_t), k_{astm} , k_{aSI} , k_{aLI} , k_{eLI} , k_{max} , k_{min}), liver metabolic rates (v_{max} , K_m , K_{met}), skin absorption parameters (D_{sc} , D_{ve} , D_{hf} , k_{aform} , k_{ahf} , PC_{sc} , PC_{scve} , PC_{hf}), blood-to-plasma ratio and tissue-to-blood partition coefficients: PC_{liv} , PC_{adp} , PC_{ppt} , PC_{hpt} , PC_{kid} , PC_{lng} , PC_{git} , PC_{skin} .

The higher the absolute sensitivity value, the more important is the parameter. These sensitivity functions are collapsed into summary values ($L1$ and $L2$ are used as selection criteria) (Eq. (2)). Based on the sensitivity functions of blood AUC to selected parameters, the identifiability of a set of parameters to be fine-tuned by calibration is then calculated. As a rule of thumb, a collinearity value less than about 20 means “identifiable” (in general, when the collinearity index exceeds 20, the linear dependence is assumed to be critical (Eq. (3)) (Brun et al., 2001; Omlin et al., 2001).

The collinearity is a measure of approximate linear dependence between sets of parameters. The higher its value, the more the parameters are related. In this context, “related” means that several parameter combinations may produce similar values of the output variables.

Monte Carlo simulations were used to quantify the impact of variability and uncertainty in parameter distributions separately by drawing parameter values according to a predefined distribution (normally distributed random samples), running the model with each of these parameter combinations, and calculating the values of the selected output variables at each output interval.

Optimization was carried out by applying the Levenberg–Marquardt algorithm to nonlinear data fitting (Moré, 1978).

Evaluation of the goodness-of-fit of the PBK models

To compare the performance of the PBK models, the coefficient of determination (R^2), mean squared error (MSE) and Akaike’s Information Criterion (AIC) (Kletting and Glatting, 2009) were used. For model discrimination, AIC penalizes models with a large number of parameters when having the same agreement in terms of mean squared error. The model with the lowest algebraic AIC therefore indicates the best agreement with experimental data while minimizing the risk of over fitting. In addition, a corrected AIC (AICc) was applied to adjust for a small number of data points (Kletting and Glatting,

2009).

Calibration and Validation

The summary of all the literature experimental data (*in vitro*, *in vivo*) used for calibrating and validating the PBK-D models for human are presented in **Tables 4-6** for all the 9 case study compounds.

Table 4: Summary of experimental data sources for toxicokinetic modelling.

Compound	<i>in vitro</i> data	<i>in vivo</i> data
Coumarin	Skin permeation: (Ritschel et al., 1977) Liver metabolism: (Born et al., 2000; Rietjens et al., 2008; Vassallo et al., 2004)	Oral data: Single: (Ritschel et al., 1979, 1977) Dermal data: Single: (Ford et al., 2001)
Estragole	Metabolic rates in liver microsomes: (Punt et al., 2009)	Oral data: (Zeller et al., 2009)
Hydroquinone	Liver metabolism: Isolated hepatocytes (Poet et al., 2004)	Oral data: (Corley et al., 2000) Dermal data: (Wester et al., 1998)
Caffeine	Skin permeation: Cosmetics Europe JRC (Mennecozzi et al., 2011) and (Doucet et al., 1998; Treffel et al., 1993) (Wilkinson et al., 2006) Liver metabolism: (Ha et al., 1996)	Oral data: Single: gelatin capsule (Lelo et al., 1986) Single: commercial dietary supplement (Csajka et al., 2005) Single: gelatin capsule (Newton et al., 1981) Single: beverage (Robertson et al., 1981) Single and repeated: gelatin capsule (Acheson et al., 1980) Repeated: coffee (Denaro et al., 1991) Dermal data : Single: Caffeine in 30%ethanol/70% propylene glycol (Otberg et al., 2008)
Ethanol	-	Oral data: (Jones et al., 1988; Wilkinson et al., 1977a, 1977b) (Lostia et al., 2013) + King's College Dermal data: (Kramer et al., 2007) Inhalation: (Nadeau et al., 2003)
Isopropanol	-	Inhalation data: (Kumagai et al., 1999) (Clewell Iii et al., 2001) Oral data: (Clewell Iii et al., 2001; Lacouture et al., 1989; Monaghan et al., 1995) Dermal data: (Turner et al., 2004)
Methyl Iodide	-	Inhalation data: (Morgan et al., 1967; Morgan and Morgan, 1967) i.v. data (Robertz-Vaupel et al., 1991)
Styrene	-	Inhalation data: (Johanson et al., 2000; Ramsey and Young, 1978; Wigaeus et al., 1984)

Nicotine	-	IV infusion: Repeated (Porchet et al., 1988) Oral data: Nicotine capsules: single (Green et al., 1999) Nicotine with the straw: single and repeated (D'Orlando and Fox, 2004) Dermal data: Nicotine patch: single and repeated (Bannon et al., 1989) Nicotine patch: single and repeated (Gupta et al., 1993) Via straw: Inhalation: repeated (Teeguarden et al., 2013)
-----------------	---	--

Table 5: Summary of experimental data sources for toxicodynamic modelling

Compound		References
Nicotine	Blood pressure Heart rate	(Fattinger et al., 1997; Mündel and Jones, 2006; Porchet et al., 1988; Sofuoglu et al., 2012)
Caffeine		(Damirchi et al., 2009; Daniels et al., 1998; Denaro et al., 1991; Karatzis et al., 2005; Mousavi et al., 2011; Ping et al., 2010; Shi, 1993; Sung et al., 1990)

Table 6: Uncertainty in model evaluation

Assumption	Uncertainty	Potential consequence for internal dose metrics	Potential risk assessment consequences
First order oral absorption	First order oral absorption has been stated to be applicable to many pharmaceuticals. However, for non-pharmaceuticals, this is uncertain. Also, the value for the first order absorption rate constant used may change for significantly different exposure conditions.	Predicted AUC as well as C_{max} values may deviate significantly from the real AUC and C_{max} (rat, human).	Significant over- or underestimation of actual risk.
No human first pass metabolism in skin	First pass metabolism reduces the internal exposure to the parent chemical.	Internal exposure to the parent chemical is overestimated (either as AUC or as C_{max}).	No consequence in comparison to default approach (just using total external dose). Larger risks predicted compared to when first pass metabolism in skin would have been taken into account.
No genetic polymorphism in the human population	The values used for V_{max} and K_m may not be valid for whole human population (it is known that there is genetic polymorphism)	Predicted AUC and V_{max} are not valid for whole human population.	Risks maybe over- or under predicted for part of the human population.

4. Discussion

The PBK models were developed to predict the internal dose metrics (AUC and C_{max}) for selected chemicals. The human models were set up to mimic typical/realistic single or repeated oral exposure (gavage as was actually used in the volunteer study for coumarin and gavage as a surrogate for the

gelatine capsule containing caffeine that was administered) and dermal human exposure scenarios. The use of GI and dermal sub-compartments in the PBK model gives a better simulation of absorption for the cosmetic ingredients than a simple model structure. However, the parameters estimated for pure substances may not apply to mixtures. For the substances studied, QSPR predictions provide a reliable alternative to in vivo dermal penetration experiments, which can be used as reliable input parameters to the dermal PBK model. Interspecies extrapolation by PBK modelling gives different results compared with a simple scaling to the human equivalent dose, which takes into account body weight and surface area only. When the same absolute dose is applied orally and dermally, different concentration–time profiles in blood and liver are obtained.

Ethical issues and considerations for 3R's impact

This study illustrates a strategy for the oral-to-dermal extrapolation of toxicity data based on PBK modelling using cosmetic ingredients. This modelling framework could be used to guide the generation of additional experimental data to refine the PBK models and extend their applicability, with the ultimate aim of applying the models in chemical risk assessment without the use of any animal studies

Known strengths and/or limitations of the method

PBK modeling requires different type of data, ranging from the physiological, biochemical, and physicochemical processes that occur in biological systems. Furthermore different age groups or certain physiological and pathological conditions can influence a chemical reaction within the body and thus should be included in the model prediction. These type of data are not always available from one source, which may lead to some confusion and to a problem in establishing a reliable source of accurate and consistent information. the JRC PBK models are also built and calibrated against the available in vivo data found in the literature and for certain type of exposure scenarios, giving higher uncertainty concerning the applicability of the model for the overall population and for different exposure scenarios. As reported in Khalil and Laer (2011) PBK models reflect the current scientific knowledge, and while some processes are known to be well characterized, others are partly or poorly characterized, as information gaps may exist. Information gaps or poor characterization of some physiological processes. Therefore, it is important to emphasize that the validity and quality of the simulations depends on the corresponding model and its incorporated data as well as its purpose, and uncertainty concerning the data used to build the model should be properly reflected.

Potential for future development

- To include Monte Carlo Simulation to take into account variability within the population. To be able to use the model and scale it up to all population.
- Develop more refined models to overcome uncertainties.

Additional Considerations

- The KNIME WEB version (KNIME WebPortal) will make the PBK model an easy to use tool.
<http://cosmospace.cosmostox.eu/app/home>

5. Contact Details

Dr. Andrew Worth

Chemicals Safety and Alternative Methods Unit - European Union Reference Laboratory for Alternatives to Animal Testing (EURL-ECVAM)
European Commission - Joint Research Centre
Corresponding author
via E. Fermi 1
Ispra (VA) 21027
email: andrew.worth@ec.europa.eu
telephone: +390332783527

Dr. Alicia Paini

Chemicals Safety and Alternative Methods Unit - European Union Reference Laboratory for Alternatives to Animal Testing (EURL-ECVAM)

European Commission - Joint Research Centre
 Method developer / user
 via E. Fermi 1
 Ispra (VA) 21027
 telephone: 39 0332 7813527

Monika Gajewska

Chemicals Safety and Alternative Methods Unit - European Union Reference Laboratory for Alternatives to Animal Testing (EURL-ECVAM)
 European Commission - Joint Research Centre
 Method developer / user
 via E. Fermi 1
 Ispra (VA) 21027
 telephone: 39 0332 7813527

Ver n°	Date	Name	Type of change
1	15 January 2015	Dr. Alicia Paini	Method summary created

6. Abbreviations & Definitions

A	amount
ADME	Absorption, Distribution, Metabolism, and Excretion
adp	adipose tissue
AIC	Akaike's Information Criterion
AICc	corrected AIC
A _{on}	amount of a chemical in organ/tissue [mg]
Area	application Area on skin [cm ²]
art	arterial blood
AUC	Area under Curve, reflects that total amount of chemical to which an organ is exposed to.
BW	body weight
C	concentration
CLR	renal clearance rate [L/h]
C _{max}	maximum concentration of a chemical in an organ
Con or Corg	concentration of a chemical in organ/tissue [mg/L]
COSMOS	EU integrated project: COSMOS develops an integrated suite of computational tools to predict the effects of long-term exposure to chemicals in humans based on in silico calculations and estimation techniques. Website: http://www.cosmos-tox.eu/
D _{hf}	diffusion coefficient in coefficient in hair follicles [cm ² /h]
Diss	dissolution from a coated matrix [L/h]
D _{sc}	diffusion coefficient in stratum corneum [cm ² /h]
D _t	drinking rate [L/h]
D _{ve}	diffusion coefficient in viable epidermis [cm ² /h]
f _{crd}	cardiac output [L/h]
GI	Gastrointestinal
GIT	Gastrointestinal tract
hpt	highly perfused tissues
HTS	High Throughput Screening
iv	intravenous

kaform	intake rate of chemical from formulation by stratum corneum [mL/h]
kahf	intake rate of a chemical from formulation by hair follicles [mL/h]
kaLI	absorption rate from large intestine lumen [1/h]
kaSI	absorption rate from small intestine lumen [1/h]
kastm	absorption rate into stomach tissue [1/h]
keLI	elimination rate in large intestine lumen [1/h]
kid	kidney
K _m	chemical concentration at which the reaction rate is half of the maximal [mg/L]
K _{met}	first order rate of formation of metabolites [L/h]
k _{min} , k _{max}	kinetic constants of stomach emptying rate of a chemical to small intestine [1/h]
KNIME	Konstanz Information Miner, is an open source data analytics, reporting and integration platform, which integrates various components for machine learning and data mining through its modular data pipelining concept. Website: www.knime.org
liv	liver
Ing	lungs
L	Skin thickness [cm]
L _{sc}	thickness of stratum corneum [cm]
L _{ve}	thickness of viable epidermis [cm]
MET _{liv}	metabolism eqations
MSE	mean squared error
ODE	Ordinary Differential Equation
on or org	organ name
PBD	Physiologically Based Dynamic (model)
PBK	Physiologically Based Kinetic (model)
PBTK	Physiologically-Based Toxicokinetic (model)
PC	partition coefficient
PC _{adp}	Tissue to blood partition coefficient - Adipose tissue
PC _{git}	Tissue to blood partition coefficient - GI tract
PC _{hf}	partition coefficient hair follicle /solvent
PC _{hpt}	Tissue to blood partition coefficient - highly perfused tissues
PC _{kid}	Tissue to blood partition coefficient - Kindney
PC _{liv}	Tissue to blood partition coefficient - Liver
PC _{Ing}	Tissue to blood partition coefficient - Lungs
PC _{org} or PC _{Con}	tissue-to- blood partition coefficients
PC _{ppt}	Tissue to blood partition coefficients - Poorly perfused tissues
PC _{sc}	partition coefficient stratum corneum/ solvent
PC _{scve}	partition coefficient stratum corneum/ viable epidermis
PC _{skn}	Tissue to blood partition coefficient - skin
PDE	Partial Differential Equation
P _{oct}	Octanol/water partition coefficient
PoD	Point of Departure
ppt	poorly perfused tissues
QSPR	quantitative structure-property relationship
R	R is a programming language and software environment for statistical computing and visualisaton. Website: http://www.r-project.org/
R ²	coefficient of determination

RPB	Plasma to blood concentration ratio
SEURAT 1	EU integrated projects cluster. SEURAT-1 Research Initiative develops a long-term research strategy leading to pathway-based human safety assessments in the field of repeated dose systemic toxicity testing of chemicals. Website: http://www.seurat-1.eu/
SF	scaling factor
skn	skin
tlag	the time lag of absorption in the skin
V	volume
ven	venous blood
V _{max}	the maximum rate metabolic rate at maximum (saturating) concentration of a chemical [mg/h]
V _{on}	volume of organ/tissue

7. Bibliography

Method development:

- Acheson, K.J., Zahorska-Markiewicz, B., Pittet, P., Anantharaman, K., Jéquier, E., (1980)
Caffeine and coffee: Their influence on metabolic rate and substrate utilization in normal weight and obese individuals.
American Journal of Clinical Nutrition 33, 989–997
- Bannon, Y.B., Corish, J., Corrigan, O.I., Devane, J.G., Kavanagh, M., Mulligan, S., (1989)
Transdermal delivery of nicotine in normal human volunteers: A single dose and multiple dose study.
European Journal of Clinical Pharmacology 37, 285–290
- Born, S.L., Caudill, D., Smith, B.J., Lehman-McKeeman, L.D., (2000)
In vitro kinetics of coumarin 3,4-epoxidation: Application to species differences in toxicity and carcinogenicity.
Toxicological Sciences 58, 23-31
- Brown, R.P., Delp, M.D., Lindstedt, S.L., Rhomberg, L.R., Beliles, R.P., (1997)
Physiological parameter values for physiologically based pharmacokinetic models.
Toxicology and Industrial Health 13, 407-484
- Brun, R., Reichert, P., Künsch, H.R., (2001)
Practical identifiability analysis of large environmental simulation models.
Water Resources Research 37, 1015-1030
- Clewell Iii, H.J., Gentry, P.R., Gearhart, J.M., Covington, T.R., Banton, M.I., Andersen, M.E., (2001)
Development of a physiologically based pharmacokinetic model of isopropanol and its metabolite acetone.
Toxicological Sciences 63, 160-172
- Corley, R.A., English, J.C., Hill, T.S., Fiorica, L.A., Morgott, D.A., (2000)
Development of a physiologically based pharmacokinetic model for hydroquinone.
Toxicology and Applied Pharmacology 165, 163-174
- Csajka, C., Haller, C.A., Benowitz, N.L., Verotta, D., 2005 (2005)
Mechanistic pharmacokinetic modelling of ephedrine, norephedrine and caffeine in healthy subjects.
British Journal of Clinical Pharmacology 59, 335-345
- Damirchi, A., Rahmani-Nia, F., Mirzaie, B., Hasan-Nia, S., Ebrahimi, M., (2009)
Effect of caffeine on metabolic and cardiovascular responses to submaximal exercise in lean and obese men.
Biomedical Human Kinetics 1, 31-35
- Daniels, J.W., Molé, P.A., Shaffrath, J.D., Stebbins, C.L., (1998)
Effects of caffeine on blood pressure, heart rate, and forearm blood flow during dynamic leg exercise.
Journal of Applied Physiology 85, 154-159

- Denaro, C.P., Brown, C.R., Jacob III, P., Benowitz, N.L., (1991)
Effects of caffeine with repeated dosing.
European Journal of Clinical Pharmacology 40, 273-278
- Doucet, O., Ferrero, L., Garcia, N., Zastrow, L., (1998)
O/W emulsion and W/O/W multiple emulsion: Physical characterization and skin pharmacokinetic comparison in the delivery process of caffeine.
International Journal of Cosmetic Science 20, 283-295
- D'Orlando, K.J., Fox, B.S., (2004)
Tolerability and pharmacokinetics of single and repeated doses of nicotine with The Straw, a novel nicotine replacement product.
Nicotine & Tobacco Research 6, 63-70
- Fattinger, K., Verotta, D., Benowitz, N.L., (1997)
Pharmacodynamics of acute tolerance to multiple nicotinic effects in humans.
Journal of Pharmacology and Experimental Therapeutics 281, 1238-1246
- Ford, R.A., Hawkins, D.R., Mayo, B.C., Api, A.M., (2001)
The in vivo dermal absorption and metabolism of [4-14C]coumarin by rats and by human volunteers under simulated conditions of use in fragrances.
Food and Chemical Toxicology 39, 153-162
- Gajewska, M., Paini, A., Sala Benito, J.V., Burton, J., Worth, A., Urani, C., Briesen, H., and Schramm, K.-W. (2015)
In vitro-to-in vivo correlation of the skin penetration, liver clearance and hepatotoxicity of caffeine.
Food and Chemical Toxicology 75, 39-49
- Gajewska, M., Worth, A., Urani, C., Briesen, H., and Schramm, K.-W. (2014)
The acute effects of daily nicotine intake on heart rate – A toxicokinetic and toxicodynamic modelling study.
Regulatory Toxicology and Pharmacology 70, 312-324
- Gajewska, M., Worth, A., Urani, C., Briesen, H., and Schramm, K.-W. (2014)
Application of physiologically-based toxicokinetic modelling in oral-to-dermal extrapolation of threshold doses of cosmetic ingredients.
Toxicology Letters 227, 189-202
- Green, J.T., Evans, B.K., Rhodes, J., Thomas, G.A.O., Ranshaw, C., Feyerabend, C., Russell, M.A.H., (1999)
An oral formulation of nicotine for release and absorption in the colon: Its development and pharmacokinetics.
British Journal of Clinical Pharmacology 48, 485-493
- Gupta, S.K., Benowitz, N.L., Jacob III, P., Rolf, C.N., Gorsline, J. (1993)
Bioavailability and adsorption kinetics of nicotine following application of a transdermal system.
British Journal of Clinical Pharmacology 36, 221-227
- Ha, H.R., Chen, J., Krähenbühl, S., Follath, F. (1996)
Biotransformation of caffeine by cDNA-expressed human cytochromes P-450.
British Journal of Clinical Pharmacology 49, 309-315
- Johanson, G., Ernstgard, L., Gullstrand, E., Lof, A., Osterman-Golkar, S., Williams, C.C., Sumner, S.C.J., (2000)
Styrene oxide in blood, hemoglobin adducts, and urinary metabolites in human volunteers exposed to 13C8-styrene vapors.
Toxicology and Applied Pharmacology 168, 36-49
- Jones, A.W., Neiman, J., Hillbom, M., (1988)
Concentration-time profiles of ethanol and acetaldehyde in human volunteers treated with the alcohol-sensitizing drug, calcium carbimide.
British Journal of Clinical Pharmacology 25, 213-221
- Karatzis, E., Papaioannou, T.G., Aznaouridis, K., Karatzi, K., Stamatelopoulos, K., Zampelas, A., Papamichael, C., Lekakis, J., Mavrikakis, M., (2005)
Acute effects of caffeine on blood pressure and wave reflections in healthy subjects: Should we consider monitoring central blood pressure?
International Journal of Cardiology 98, 425-430
- Khalil F., and Stephanie Læer S., (2011)
Physiologically based pharmacokinetic modeling: methodology, applications, and limitations with a focus on its role in pediatric drug development

Journal of Biomedicine and Biotechnology 1, 13

- Kletting, P., Glatting, (2009)
Model selection for time-activity curves: The corrected Akaike information criterion and the F-test.
Zeitschrift für Medizinische Physik 19, 200-206
- Kramer, A., Below, H., Bieber, N., Kampf, G., Toma, C.D., Huebner, N.-O., Assadian, O., (2007)
Quantity of ethanol absorption after excessive hand disinfection using three commercially available hand rubs is minimal and below toxic levels for humans.
BMC Infectious Diseases 7
- Kumagai, S., Matsunaga, I., (1995)
Physiologically based pharmacokinetic model for acetone.
Occupational and Environmental Medicine 52, 344-352
- Lacouture, P.G., Heldreth, D.D., Shannon, M., Lovejoy Jr., (1989)
The generation of acetonemia/acetonuria following ingestion of a subtoxic dose isopropyl alcohol.
American Journal of Emergency Medicine. 7, 38-40
- Lelo, A., Birkett, D.J., Robson, R.A., Miners, J.O., (1986)
Comparative pharmacokinetics of caffeine and its primary demethylated metabolites paraxanthine, theobromine and theophylline in man.
British Journal of Clinical Pharmacology 22, 177-182
- Lostia, A.M., Vicente, J.L., Cowan, D.A., (2013)
Measurement of ethyl glucuronide, ethyl sulphate and their ratio in the urine and serum of healthy volunteers after two doses of alcohol.
Alcohol and Alcoholism 48, 74-82
- Mennecozzi, M., Landesmann, B., Harris, G.A., Liska, R., Whelan, M., (2011)
Hepatotoxicity Screening Taking a Mode-Of-Action Approach Using HepaRG Cells and HCA.
Proceedings WC08.
Alternatives to Animal Experimentation (ALTEX) 1/12
- Monaghan, M.S., Olsen, K.M., Ackerman, B.H., Fuller, G.L., Porter, W.H., Pappas, A.A., (1995)
Measurement of serum isopropanol and the acetone metabolite by proton nuclear magnetic resonance: Application to pharmacokinetic evaluation in a simulated overdose model.
Journal of Toxicology - Clinical toxicology. 33, 141-149
- Morgan, A., Morgan, D.J., Evans, J.C., Lister, B.A., (1967)
Studies on the retention and metabolism of inhaled methyl iodide. II. Metabolism of methyl iodide.
Health Physics 13, 1067-1074
- Morgan, D.J., Morgan, A., (1967)
Studies on the retention and metabolism of inhaled methyl iodide. I. Retention of inhaled methyl iodide.
Health Physics 13, 1055-1065
- Moré, J.J., (1978)
The Levenberg-Marquardt algorithm: implementation and theory. Lect. Notes Math.
630, 105-116
- Mousavi, A., Koushki Jahromi, M., Salesi, M., Daryanoush, F., Khoshnam, E., A, N., Hemati, M., (2011)
Impact of caffeine on heart rate and blood pressure at rest and during exercise. Jahrom Univ. Med. Sci.
9
- Mündel, T., Jones, D.A., (2006)
Effect of transdermal nicotine administration on exercise endurance in men.
Experimental physiology 91, 705-713
- Nadeau, V., Lamoureux, D., Beuter, A., Charbonneau, M., Tardif, R., (2003)
Neuromotor effects of acute ethanol inhalation exposure in humans: A preliminary study.
Journal of Occupational Health 45, 215-222
- Newton, R., Broughton, L.J., Lind, M.J., Morrison, P.J., Rogers, H.J., Bradbrook, I.D., (1981)
Plasma and salivary pharmacokinetics of caffeine in man.
European Journal of Clinical Pharmacology 21, 45-52
- Omlin, M., Brun, R., Reichert, P., (2001)
Biogeochemical model of Lake Zürich: Sensitivity, identifiability and uncertainty analysis.
Ecological Modelling 141, 105-123
- Omlin, M., Reichert, P., Forster, R (2001)

Biogeochemical model of Lake Zürich: Model equations and results.

Ecological Modelling 141, 77-103

- Otberg, N., Patzelt, A., Rasulev, U., Hagemeister, T., Linscheid, M., Sinkgraven, R., Sterry, W., Lademann, J., (2008)
The role of hair follicles in the percutaneous absorption of caffeine.
British Journal of Clinical Pharmacology 65, 488-492
- Ping, W.C., Keong, C.C., Bandyopadhyay, A., (2010)
Effects of acute supplementation of caffeine on cardiorespiratory responses during endurance running in a hot & humid climate.
Indian Journal of Medical Research 132, 36-41
- Poet, T.S., Carlton, B.D., Deyo, J.A., Hinderliter, P.M., (2010)
Hydroquinone PBPK model refinement and application to dermal exposure.
Food and Chemical Toxicology 48, 3085-3092
- Poet, T.S., Wu, H., English, J.C., Corley, R.A., (2004)
Metabolic rate constants for hydroquinone in F344 rat and human liver isolated hepatocytes: Application to a PBPK model.
Toxicological Sciences 82, 9-25
- Porchet, H.C., Benowitz, N.L., Sheiner, L.B., (1988)
Pharmacodynamic model of tolerance: Application to nicotine.
Journal of Pharmacology and Experimental Therapeutics 244, 231-236
- Punt, A., Paini, A., Boersma, M.G., Freidig, A.P., Delatour, T., Scholz, G., Schilter, B., van Bladeren, P.J., Rietjens, I.M.C.M., (2009)
Use of physiologically based biokinetic (PBBK) modeling to study estragole bioactivation and detoxification in humans as compared with male rats.
Toxicological Sciences 110, 255-269
- Ramsey, J.C., Young, J.D., (1978)
Pharmacokinetics of inhaled styrene in rats and humans.
Scandinavian Journal of Work, Environment and Health 4, 84-91
- Rietjens, I.M.C.M., Boersma, M.G., Zaleska, M., Punt, A., (2008)
Differences in simulated liver concentrations of toxic coumarin metabolites in rats and different human populations evaluated through physiologically based biokinetic (PBBK) modeling.
Toxicology In Vitro 22, 1890-1901
- Ritschel, W.A., Brady, M.E., Tan, H.S.I., (1979)
First-pass effect of coumarin in man. *International Journal of Clinical pharmacology, Therapy and Toxicology*. 17, 99-103
- Ritschel, W.A., Brady, M.E., Tan, H.S.I., Hoffmann, K.A., Yiu, I.M., Grummich, K.W., (1977)
Pharmacokinetics of coumarin and its 7-hydroxy-metabolites upon intravenous and peroral administration of coumarin in man.
European Journal of Clinical Pharmacology 12, 457-461
- Robertson, D., Wade, D., Workman, R., Woosley, R.L., Oates, J.A., (1981)
Tolerance to the humoral and hemodynamic effects of caffeine in man.
Journal of Clinical Investigation 67, 1111-1117
- Robertz-Vaupel, G.-M., Bierl, R., Von Unruh, G., (1991)
Intravenous methyl iodide intoxication - detoxication by haemoperfusion [INTRAVENÖSE METHYLJODIDINTOXIKATION - DETOXIKATION DURCH HAMOPERFUSION]. *Anesthesiol. Intensivmed. Notfallmedizin Schmerztherapie* 26, 44-47
- Shi, J. (1993)
Pharmacokinetic-pharmacodynamic modeling of caffeine: Tolerance to pressor effects.
Clinical Pharmacology and Therapeutics 53, 6-14
- Soetaert, K., (2010)
R package FME: inverse modelling, sensitivity, Monte Carlo - applied to a dynamic simulation model.
- Soetaert, K., Petzoldt, T., (2010)
Inverse Modelling, Sensitivity and Monte Carlo Analysis in R Using Package FME. *J. Stat. Softw.* 33, 1-28

- Sofuoglu, M., Herman, A.I., Nadim, H., Jatlow, P., (2012)
Rapid nicotine clearance is associated with greater reward and heart rate increases from intravenous nicotine.
Neuropsychopharmacology 37, 1509-1516
- Sung, B.H., Lovallo, W.R., Pincomb, G.A., Wilson, M.F (1990)
Effects of caffeine on blood pressure response during exercise in normotensive healthy young men.
American Journal of Cardiology 65, 909-913
- Teeguarden, J.G., Housand, C.J., Smith, J.N., Hinderliter, P.M., Gunawan, R., Timchalk, C.A (2013)
A multi-route model of nicotine-cotinine pharmacokinetics, pharmacodynamics and brain nicotinic acetylcholine receptor binding in humans.
65, 12–28
- Treffel, P., Panisset, F., Humbert, P., Remoussenard, O., Bechtel, Y., Agache, P., (1993)
Effect of pressure on in vitro percutaneous absorption of caffeine
Acta Dermato-Venereologica 73, 200-202
- Turner, P., Saeed, B., Kelsey, M.C., (2004)
Dermal absorption of isopropyl alcohol from a commercial hand rub: Implications for its use in hand decontamination.
Journal of Hospital Infection 56, 287-290
- Vassallo, J.D., Hicks, S.M., Daston, G.P., Lehman-McKeeman, L.D., (2004)
Metabolic detoxification determines species differences in coumarin-induced hepatotoxicity.
Toxicological Sciences 80, 249-257
- Wester, R.C., Melendres, J., Hui, X., Cox, R., Serranzana, S., Zhai, H., Quan, D., Maibach, H.I., (1998)
Human in vivo and in vitro hydroquinone topical bioavailability, metabolism, and disposition
Journal of Toxicology and Environmental Health - Part A 54, 301-317
- Wigaeus, E., Lof, A., Byfalt Nordqvist, M., (1984)
Uptake, distribution, metabolism, and elimination of styrene in man. A comparison between single exposure and co-exposure with acetone. *Br. J. Ind. Med*
41, 539-546
- Wilkinson, P.K., Sedman, A.J., Sakmar, E., Kay, D.R., Wagner, J.G., (1977)
Pharmacokinetics of ethanol after oral administration in the fasting state.
Journal of Pharmacokinetics and Biopharmaceutics 5, 207-224
- Wilkinson, P.K., Sedman, A.J., Sakmar, E., Lin, Y.-J., Wagner, J.G., (1977)
Fasting and nonfasting blood ethanol concentrations following repeated oral administration of ethanol to one adult male subject.
Journal of Pharmacokinetics and Biopharmaceutics 5, 41-52
- Wilkinson, S.C., Maas, W.J.M., Nielsen, J.B., Greaves, L.C., van de Sandt, J.J.M., Williams, F.M., (2006)
Interactions of skin thickness and physicochemical properties of test compounds in percutaneous penetration studies.
Chemosphere 79, 405-413
- Zeller, A., Horst, K., Rychlik, M., (2009)
Study of the metabolism of estragole in humans consuming fennel tea.
Chemical Research in Toxicology 22, 1929-1937

General references:

- Adler, S.; Basketter, D.; Creton, S.; Pelkonen, O.; van Benthem, J.; Zuang, V.; Andersen, K. E.; Angers-Loustau, A.; Aptula, A.; Bal-Price, A.; Benfenati, E.; Bernauer, U.; Bessems, J.; Bois, F. Y.; Boobis, A.; Brandon, E.; Bremer, S.; Broschard, T.; Casati, S.; Coecke, S.; Corvi, R.; Cronin, M.; Daston, G.; Dekant, W.; Felter, S.; Grignard, E.; Gundert-Remy, U.; Heinonen, T.; Kimber, I.; Kleinjans, J.; Komulainen, H.; Kreiling, R.; Kreysa, J.; Batista Leite, S.; Loizou, G.; Maxwell, G.; Mazzatorta, P.; Munn, S.; Pfuhler, S.; Phrakonkham, P.; Piersma, A.; Poth, A.; Prieto, P.; Repetto, G.; Rogiers, V.; Schoeters, G.; Schwarz, M.; Serafimova, R.; Tahti, R.; Testai, E.; van Delft, J.; van Loveren, H.; Vinken, M.; Worth, A. and Zaldivar, J.M. (2011)
Alternative (non-animal) methods for cosmetics testing: current status and future prospects – 2010. [Link to final report](#) (last access 19.06.2013) [Link to key policy documents](#) (last access 19.06.2013)
Archives of Toxicology 85, 367-485

Last update: 01 April 2015

DB-ALM Method summary n° 162 : The Virtual Cell Based Assay

Systemic Toxicity

The **Virtual Cell Based Assay (VCBA)** is a mathematical model for predicting the fate of a chemical within an *in vitro* system. The model can be used to simulate intracellular concentrations of selected chemicals in cell lines, and it can be coupled with physiologically based kinetic (PBK) models to predict human *in vivo* toxic doses from concentrations that cause effects *in vitro*.

1. General Information

Corresponding author(s):

Dr. Andrew Worth, European Commission - Joint Research Centre

Dr. Jose Vicente Sala Benito, European Commission - Joint Research Centre

Date of data sheet creation: 10 February 2015

Last version number: 1

Last update: 10 February 2015

Section **5. Contact Persons** provides a complete list of the authors and method developers.

The method summary presented here describes a mathematical model.

In the context of this document, the term "method" in the paragraph headings refers to the VCBA model.

Proprietary and/or Confidentiality Issues:

None - publicly available.

The model has been developed by the Joint Research Centre (JRC), with partial support from the FP7 COSMOS project.

Status

Development

Development of **Version 1.0** completed. The model is currently under revision. The VCBA models for simulation of single and repeat exposure are built in R and have been translated in KNIME (the Konstanz Information Miner) automated workflows, as a part of the COSMOS project (Integrated *In Silico* Models for the Prediction of Human Repeated Dose Toxicity of Cosmetics to Optimise Safety), within the SEURAT-1 Research Initiative (Safety Evaluation Ultimately Replacing Animal Testing 1) cluster (2011-2015). More information on the COSMOS project is available on its dedicated website:

<http://www.cosmostox.eu>



2. Method Definition

Purpose of the Method

TYPE OF TESTING	:	Part of an integrated testing strategy
LEVEL OF TOXICITY ASSESSMENT	:	Risk assessment
PURPOSE OF TESTING	:	Safety

Context of use

The **Virtual Cell Based Assay (VCBA)** is a mathematical model for predicting the fate of a chemical within an *in vitro* system. The model is designed to simulate intracellular concentrations of selected chemicals in cell lines, and can be coupled with physiologically based kinetic (PBK) models to predict human *in vivo* toxic doses from concentrations that cause effects *in vitro* (*in vitro* to *in vivo* extrapolation), thereby supporting chemical risk assessments.

The characterization of the concentration that produces an effect (whether this is a perturbation of a molecular pathway or an apical toxic endpoint) is necessary at two levels: first, in analysing the results of *in vitro* experiments, since “nominal” concentrations do not represent the real concentration experienced by the cell; and, second, in extrapolating to humans, since the true concentration experienced by cells within the target organ is more representative for human toxicity assessment. In order to analyse the results of *in vitro* experiments, a VCBA was developed for liver cell lines (3T3cBalbc, HepG2 and HepaRG) and it has been extended to include cell lines from different organs: lung A459 cells and from heart cardiomyocytes (Paini A., et al., and Sala Benito et al., both in preparation).

VCBA simulates processes in an *in vitro* system, especially the fate of a chemical within the well, taking into account partitioning with protein, lipids, and plastic binding (Zaldívar et al., 2010 and 2011). The VCBA consists also of a cell partitioning and cell growth model (G1, S, G2, M phases). An additional feature takes into account the cell toxicity. The latter part of the model was based on two parameters: the no-effect concentration (NEC) and the killing rate (kt), which are linked to experimental cell viability. The main simulated property is the intracellular concentration of a specific chemical within the cell, and its corresponding effect on cell viability (Zaldívar et al., 2010 and 2011). A mitochondrial compartment has been added to the VCBA. This model was introduced in order to be sure that the model predictions were within the order of magnitude of the *in vitro* results obtained.

Target Users and Uses

For in vitro toxicologists:

External dose → concentration in blood or selected organ → → → In vitro testing strategy planning

For Risk Assessors:

External dose → concentration in blood or selected organ → → → Point of Departure (PoD) for extrapolation (Risk Assessment).

Potential Applications

In particular, it is suggested that the use of VCBA model coupled to **Physiologically Based Kinetic (PBK) models [DB-ALM Method Summary No. 161]** to translate the *in vitro* to the *in vivo* kinetics of chemicals can be further used to extended to simulate the dynamics of chemical-induced effects on different type of cells, and can be used for the quantitative *in vitro* to *in vivo* extrapolation (QIVIVE).

Scientific Principle of the Method

Model components

The VCBA model has been developed by the JRC, based on the initial and original formulation of Zaldívar et al. (2010, 2011). The original model has been refined, extended, and recoded for implementation in the KNIME platform.

It consists of ordinary differential equations whose solution allows the calculation over time of the dissolved concentration of a chemical in cell culture as well as the internal concentration in the cells (**Figure 1**). The VCBA model currently comprises four or five linked models (1-5, below) which take into account the experimental set up conditions (6 below):

[1] The ***in vitro* fate and transport model** that calculates the time-dependent chemical concentration in the medium as well as in the headspace. This takes into consideration a series of processes including evaporation, partitioning of chemicals from the dissolved phase to serum proteins and lipids, adsorption onto the plastic, and also degradation and metabolism.

[2] The **cell partitioning model** that was built on the assumption that once the chemical is taken up by the cell, a partitioning occurs between three compartments: one aqueous fraction and two non-aqueous fractions corresponding to structural components (proteins) and energy resources (lipids).

[3] The **cell growth and division model** that is based on a four stage based approach, with each stage corresponding to one of the four cell cycle phases: G1, S, G2 and M.

[4] **Toxicity and effects model**. The direct effects of a chemical concentration, C , on cell dynamics (survival/mortality) are expressed by using the killing rate, kt , and the no effect concentration, NEC .

[5] **Mitochondrial compartment**. The *in vitro* effects were monitored in terms of the mitochondrial membrane potential (mmp) and were translated into a mathematical model by the addition of a new mitochondrial compartment within the cell.

[6] **Experimental set up**: takes into account the surface, area, size and shape of the well or well-plate, as well as incubation conditions (usually 37°C).

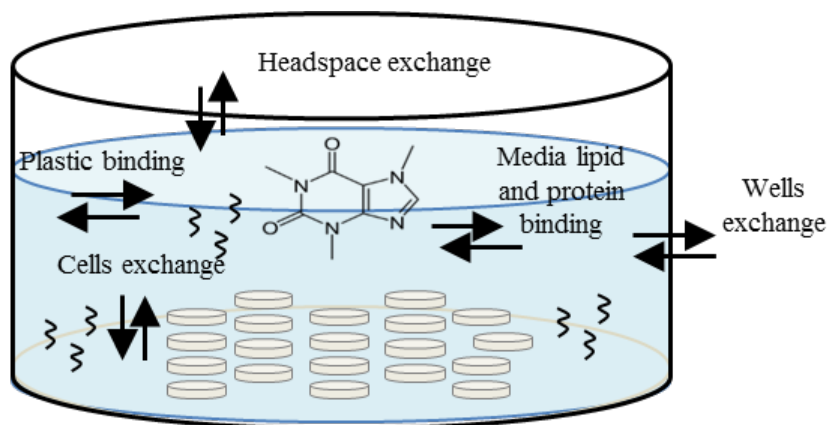


Figure 1. Schematic representation of the VCBA model.

Procedure Description

Biological Endpoint and Measurement:

CELL VIABILITY expressed as %

CELL GROWTH AND FATE

CELLULAR FUNCTIONAL PARAMETERS: INTRACELLULAR CONCENTRATION

Basic procedure:

The VCBA model is a mathematical approach. The model is coded and implemented in the open-source KNIME platform, available via COSMOS space:

- <http://cosmospace.cosmostox.eu/app/home>

The KNIME workflow is designed to be user friendly, and is divided into three zones: input, core and output (Sala Benito et al., in preparation). In the input zone, the user selects the test compound, the cell type, and the experimental conditions; in the core zone the application of the mathematical model is hidden; and the results are displayed in the output zone in graphical or tabular format.

The workflow gives us possible use of the VCBA model in an automated way and user-friendly way. This workflow and the VCBA model can be used not only to design *in vitro* experiments but also for risk assessment purposes.

Examples of the equations used in the individual model components are described in the next section:

2.1 Model description .

A presentation on the COSMOS Automated *in silico* tools for *in vitro* to *in vivo* extrapolation was recorded for the American Society for Cellular and Computational Toxicology (ASCCT). Method developer Dr. Alicia Paini explains the composition of the model and the use of the KNIME workflows:

- <https://youtu.be/vJc1UEbuWdE>

2.1 Model Description

Model Equations [Zaldivar et al., 2010, 2011]

The ***in vitro* fate and transport model** consists of a dynamic mass balance that includes a time-variable chemical transport and fate model for calculating the chemical concentration in the medium as well as in the headspace. The gas phase has been included to consider, in a second step, the possible losses and cross contamination between the 96 wells in the TC plates, since the TC are not hermetic even though the system was designed to minimize this aspect. To quantify this phenomenon, there are not enough experimental data at the moment.

Assuming a well-mixed medium and headspace and that the sorption processes are fast compared with the other processes then the mass balance equations for both compartments can be written as:

- Total concentration in the medium:

$$V_M \cdot \frac{dC_T}{dt} = A_S \cdot F_{AW} - k \cdot V_M \cdot C_{dis} \quad (1)$$

where V_M refers to the volume of the medium (m^3) and T refers to the total concentration ($mg \cdot m^{-3}$); the first term of the rhs represents the transfer of chemical across the air-water interface whereas the second term represent transformation/losses from the medium, e.g. degradation, decomposition, etc.

- Total concentration in the air:

A similar equation can be written for the headspace:

$$V_H \cdot \frac{dC_{Air}}{dt} = A_S \cdot F_{AW} - A_I \cdot F_I \quad (2)$$

To model the partitioning of an organic chemical in the medium (Kramer, 2010), we can consider that the compounds are either purely dissolved (C_{diss}), bound to the serum in the culture medium (C_S), bound to the cells (C_b) and bound to the (plastic/glass) surface of the culture vessel (C_p). Therefore, the total concentration of an organic contaminant in the medium, C_T , can be described by following equation:

$$C_T = C_{dis} + C_S + C_b + C_P \cdot \frac{S_M}{V_M} \quad (3)$$

where S_M refers to the surface of the well in contact with the medium.

In the **cell partitioning model** we extrapolate the equation concerning the cell model. The total number of moles of a compound (n_{tot}) in the cell is their sum in the different compartments (Zaldivar et al., 2010, 2011):

$$n_{tot} = n_{aq} + n_p + n_L = (V_{aq} \cdot C_{aq} + V_P \cdot C_P + V_L \cdot C_L) \quad (4)$$

where the V_i 's refer to the compartment volumes (l) and the C_i 's refer to the compartments concentration ($\text{mol} \cdot \text{l}^{-1}$). Also the total number of moles of a chemical can be expressed as:

$$n_{tot} = W \cdot C_b / MW \quad (5)$$

where W is the organism weight (g), MW is the molecular weight of the chemical ($\text{g} \cdot \text{mol}^{-1}$) and C_b is the chemical concentration in the cell ($\text{g} \cdot \text{g} \cdot \text{w}^{-1}$).

The chemical is assumed to be in equilibrium between the different compartments with fixed values partition coefficients: $K_P = C_P / C_{aq}$ and $K_L = C_L / C_{aq}$.

The time evolution of this substance in the cell can be calculated by a simple mass balance, assuming that the uptake and depuration rates r_{ad} and r_{da} ($\text{l} \cdot \text{cm}^{-2} \cdot \text{s}^{-1}$) are proportional to the surface area of the cell (passive diffusion) and the transfer occurs through the aqueous compartment only as:

$$\frac{dn_{tot}}{dt} = V^{2/3} (r_{da} \cdot A C_{diss} - r_{ad} \cdot A C_{aq}) \quad (6)$$

where C_{diss} and C_{aq} refer to the chemical concentration ($\text{mol} \cdot \text{l}^{-1}$) outside of the cell and in the aqueous compartment of the cell ($\text{mol} \cdot \text{l}^{-1}$), respectively. Applying the chain rule of derivation to equation 4 we have:

$$\frac{dn_{tot}}{dt} = \frac{1}{MW} \cdot \left(W \cdot \frac{dC_b}{dt} + C_b \cdot \frac{dW}{dt} \right) \quad (7)$$

and rearranging terms we obtain:

$$\frac{dC_b}{dt} = \frac{MW \cdot V^{2/3}}{W} \cdot (r_{da} \cdot C_{diss} - r_{ad} \cdot C_{aq}) - \frac{C_b}{W} \cdot \frac{dW}{dt} \quad (8)$$

the last term represents the dilution due to growth of the cell that in the case of HepaRG and cardiomyocytes can be neglected.

Since the concentration in the aqueous fraction C_{aq} is not a value that is measured, then we have to convert in terms of C_b using the partitioning approach. The wet weight, W can also be expressed as a function of the volumes of the different compartments:

$$W = \rho \cdot V = \rho \cdot (V_{aq} + V_p + V_L) \quad (9)$$

On the other hand:

$$V_p = W_p / \rho_p \quad (10)$$

$$V_L = W_L / \rho_L \quad (11)$$

$$V_{aq} = W_{aq} / \rho_{aq} \quad (12)$$

where W_p , W_L and W_{aq} are the masses of proteins, lipids and aqueous compartments in the cells and ρ_p , ρ_L and ρ_{aq} their densities.

To find the correlation between C_{aq} and C_b we have to combine n_{tot} in equations 4-5, the partition coefficients and equations 7-9, then we have:

$$C_{aq} = \frac{C_b}{MW \cdot \left(\frac{f_{aq}}{\rho_{aq}} + \frac{f_L}{\rho_L} K_L + \frac{f_p}{\rho_p} K_p \right)} \quad (13)$$

where f_i 's refer to the mass fraction of each compartment (aqueous, lipid, proteins) in the cell. Replacing this equation into equation 6 and rearranging we obtain:

$$\frac{dC_b}{dt} = \left(\frac{MW \cdot V^{2/3}}{W} r_{da} \right) C_{diss} - \left(\frac{V^{2/3}}{W \left(\frac{f_{aq}}{\rho_{aq}} + \frac{f_L}{\rho_L} K_L + \frac{f_p}{\rho_p} K_p \right)} r_{ad} \right) C_b - \left(\frac{1}{W} \frac{dW}{dt} \right) C_b \quad (14)$$

The model takes into account a **cell growth and division** which is a four steps mechanism: G1, S, G2, M. In the stage-based type of modelling the matrix **A**, called Leslie matrix, which describes the transformation of a cell population from time t to time $t+1$,

$$n_{t+1} = A \cdot n_t \quad (15)$$

has the following structure:

$$A = \begin{bmatrix} P_1 & 0 & 0 & F \\ G_1 & P_2 & 0 & 0 \\ 0 & G_2 & P_3 & 0 \\ 0 & 0 & G_3 & P_4 \end{bmatrix} \quad (16)$$

where n_t is a vector describing the cell population at each stage at time t , P_i is the probability of surviving and staying in stage i , G_i is the probability of surviving and growing into the next stage and F is the fecundity rate per unit time (h), $i=1,2,3,4$. Both P_i and G_i are functions of the survival probability p_i and grow probability γ :

$$P_i = p_i \cdot (1 - \gamma) \quad (17)$$

$$G_i = p_i \cdot \gamma_i \quad (18)$$

where

$$p_i = \exp(-z_i) \quad (18)$$

and

$$\gamma_i = \frac{(1-p_i) \cdot p_i^{d_i-1}}{1-p_i^{d_i-1}} \quad (19)$$

where z_i is the hourly instantaneously mortality rate and d_i is the duration (h) within the i -th stage. To study the direct effects of a chemical concentration, C , on survival [**Toxicity and effects model**] may be expressed using the following equation:

$$p_i = \exp(-z_i) \quad (20)$$

by the addition of a term which can be written as:

$$z_i = \frac{z_i + kt(C_b - NEC) \dots C_b > NEC}{z_i} \quad (21)$$

where C_b is the internal concentration of the toxicant in the cell, kt is the killing rate and NEC is the no effect concentration term.

The next differential equation describes the evolution in time of the cell connected to the toxicodynamic effect described by the two parameters kt and NEC :

$$\frac{dn}{dt} = -kt \cdot (C_b - NEC) \cdot n \quad (22)$$

The Mitochondrial model

The refined cell model is described by four compartments: lipid, proteic, aqueous and mitochondrial (Figure 2). The interchange of the chemical through the cell membrane occurs by a diffusion mechanism and then the chemical is distributed into the 3 compartments of the cell (aqueous, lipidic and proteic by partitioning) and depending on the chemical size by passive (<1500Da) or mediated (>5000Da) diffusion into mitochondrial (Alberts et al., 2002).

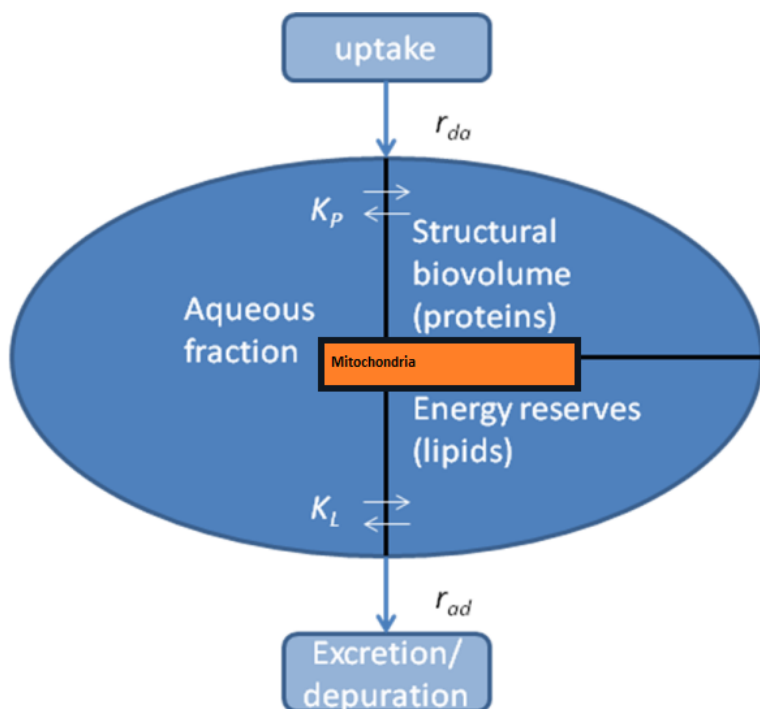


Figure 2: A new four-compartment cell model, including a single mitochondrial compartment.

With regard to mitochondrial model compartment, the chemical enters through a diffusive mechanism from the aqueous department to mitochondria:

$$\frac{dC_{mito}}{dt} = K_{mit} \cdot (C_{aq} - C_{mit}) \quad (23)$$

Where C_{aq} is the concentration in the aqueous phase, C_{mit} is the concentration inside of the mitochondria and K_{mit} is a diffusive parameter depending on the chemical and cell line applied.

Finally, solving together the differential equations the model can simulate the predictions of the viability, intracellular concentration, and concentration inside of mitochondria and mitochondria membrane potential can be obtained.

Experimental set up:

The total volume of the well is given by assuming a truncated cone equation:

$$V_W = 1/3 (r^2 + r \cdot R + R^2) \cdot h \text{ [m}^3\text{]} \quad (25)$$

The headspace volume [m³] is given by subtracting from equation 25 the Volume of the medium.

The surface of the well in contact with the medium, $S = p \cdot (r+r)g + S$,

where r is the radius of the occupied volume and g is the slant height.

Area of the cell-based assay medium, $A_s = \pi \cdot r_m^2$.

Model Input Parameters

[1] A fate model: chemical properties

For the fate model the following parameters are needed. They can be calculated with several *in silico*

tools, EPI suite, QSARToolBox etc.

1. Henry constant; 2. LogKow; 3. Air and water degradation; 4. Molar volume; 5. Atomic diffusion.

[2] A cell model: mass fraction of compartments [protein, lipids, water], partition coefficients within the cell, uptake rate. (Example with HepaRG)

A virtual cell (VC) model for the HepaRG liver cell line was established based on the VCBA model described in Zaldivar et al., (2011 and 2012). Briefly, the model takes into account the fate of a compound within a cell. The cell is divided in 3 compartments (aqueous, lipids, and proteins) a cell growth model was derived, however it was set to initial G1 phase, since the HepaRG cell line does not proliferate. Finally the model also features a toxicodynamic effect model that takes into account mortality, by means of cell viability. The mortality parameters of the cells due to chemical effect are described by the NEC and k_t .

Table 1. Examples of Parameters for the HepaRG virtual cell model (Zaldivar et al., 2010, 2011)

Parameter type	Abbreviation used in the model	Value	Units
Mass fraction of compartment f_x (aq- aqueous , l-lipids, p-proteins)	f_{aq} ,	0.72	% weight
	f_L ,	0.012	
	f_P	0.268	
Uptake rate	r_{da}	35.208	L m ⁻² h ⁻¹
Elimination rate	r_{ad}	35.208	L m ⁻² h ⁻¹
Wet weight	W	1.79•10 ⁻⁹	gr
Volume of the cell	V	1.67•10 ⁻¹⁵	m ³
Cell cycle* duration	G1	849	h
Mortality	M	1.19•10 ⁻⁶	h ⁻¹

*HepaRG cell do not proliferate the models keeps only one cell cycle step G1.

[3] A cell growth model: information on cell cycle dynamics. See Table 1. Cell cycle duration G1, S, G2, M.

[4] A toxic effect model [5] mitochondrial model: NEC and killing rate k_t

In order to optimize the model experimental curves, concentration response curve are needed. These *in vitro* curves should be measured at 24 hrs for cytotoxicity and mmp effects.

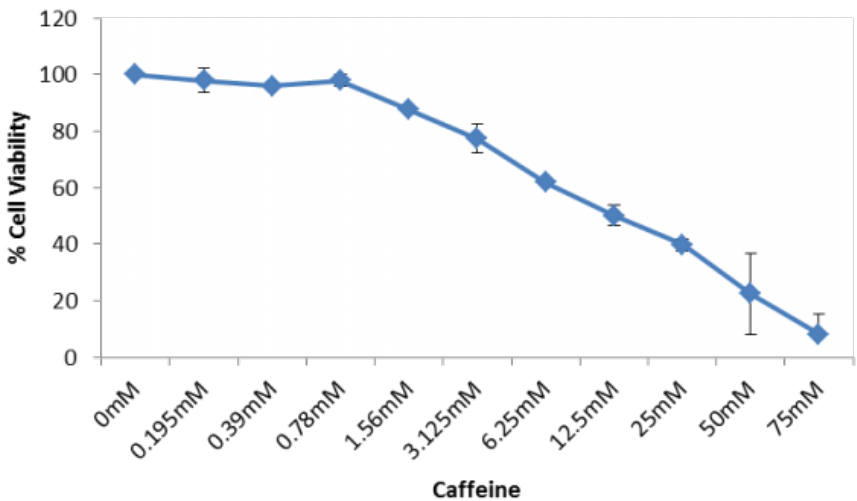


Figure 3. Example of concentration response curve, measured output cell viability.

[6] Experimental set up

List of parameters to describe the experimental set up:

- Incubator temperature

The parameters used to describe the configuration of the well test plate, are (examples for 96 Well plate):

- Top internal radius: $3.425 \cdot 10^{-3}$ m.
- Bottom internal diameter: $3.175 \cdot 10^{-3}$ m.
Depth: $10.76 \cdot 10^{-3}$ m.
- Bottom area $3.16 \cdot 10^{-5}$ m²
- Volume of medium. 100uL ($V_M = 10^{-7}$ m³),
- Headspace volume (m³): $V_H = 2.683 \cdot 10^{-7}$ m³.
- Amount of protein in medium: Assuming a 5% (v/v) serum in the medium, then $[S]_0 = 2.34 \cdot 10^{-2}$ mol protein m⁻³.

Model Settings:

The model was built taking into consideration both single and repeat exposure.

1) VCBA Model mode:

This mode is chosen to run the Cell Based Assay Model, based on a concentration-response curve. The concentrations are written in one column of an XLS file (node XLS Reader: Simulation Data). The final result is a graphical representation made up of four graphics; with the evolution in time of the chemical concentrations in the medium, headspace, and inside the cell, and the number of cells.

2) Parameter Optimization mode:

This mode is selected to optimise values of the toxicodynamic parameters, the No Effect Concentration (*NEC*) and the killing rate (*kt*). The initial values of the *NEC* and *kt*, along with experimentally derived cell viability values, are entered in the Optimization Parameters tab.

3) Check Optimization mode:

From the new values of *NEC* and *kt* obtained by running the Parameter Optimization mode and the experimental values in the XLS file, a figure can be drawn which is used to check how good the model fits the experiments.

4) Repeat dose mode:

This mode simulates the effects of different doses and regular intervals of exposure. The results are depicted as a graph, representing a time-response curve of the cell viability (using different doses).

3. Method Performance

Robustness of the Method

General performance measures:

The VCBA model has undergone an extensive evaluation, described in the next section: **3.1 Evaluation of Model Performance.**

3.1 Evaluation of Model Performance

Optimization Analysis

We briefly explain how to optimize the values of the Toxicity effect model: the No Effect Concentration (*NEC*) and the killing rate (*kt*). The initial values of the *NEC* and *kt*, along with experimentally derived cell viability values, are entered in the Optimization Parameters tab.

Table 2. Input experimental values introduced into workflow: concentration (of the test compound), viability (response), and standard deviation.

CONC (µg/mL)	VIABILITY (%)	std
0.78	104.15	7.412476
1.56	106.40	11.24928
3.13	104.08	7.671877
6.25	108.16	10.08127
8.78	100.31	14.43041
12.5	81.37	10.16028
25	58.70	9.369224
44.44	42.82	8.410667
50	32.91	6.169209
66.67	25.37	7.814317
100	16.97	5.564936

Input experimental values are written into three columns of the XLS file introduced into workflow by the XLS Reader node: concentration (of the test compound), viability (response), and standard deviation (or error), Table 2. The results of the parameter optimization are new values of NEC and kt that minimize the differences between the experimental viability and the viability from the model (error). The input parameters can be manually (recommended) or automatically selected by the user from the XLS Reader node. In the automatic mode the selection is the entire data sheet. If the selection is done manually, the user must select the rows and columns of the excel sheet which are of interest to him.

Evaluation of the goodness-of-fit of the VCBA model

From the new values of NEC and kt obtained by running the Parameter Optimization mode and the experimental values in the XLS file, a figure can be drawn which is used to check how good the model fits the experiments (goodness of fit). The core model calculates the simulated viability for each of the chemical concentrations using the NEC and kt values calculated in the Parameter Optimization Mode. This functionality was separated from the Parameter Optimization in order to operate and test the values of NEC and kt from other sources without running each time the time-consuming Parameter Optimization Mode.

Test Compounds

Chemicals used

Table 3 reports the list of chemicals present in the VCBA model and corresponding cell line tested.

Table 3. List of chemicals tested for each type of cell type

Name of Compound	Single Exposure				Repeat Exposure
	Cardiomyocyte	HepG2	HepaRG	A549	HepaRG
acetaminophen	X	X	X		X
acetylsalicylic acid		X			
acrylamide		X			
amiodarone	X	X	X		X
caffeine	X	X	X		X
carbamazepine		X			
colchicine		X			
coumarin			X		
cycloheximide		X	X		
diazepam		X			
diquat dibromide		X			

estragole		X	X
ethanol	X		
isopropanol	X		
Limonene			X
nicotine	X	X	
pentachlorophenol	X		
rifampin	X		
sodium lauryl sulphate	X	X	
tert-butylhydroperoxide	X		
tetracycline hydrochloride	X		
thioridazine hydrochloride	X		
valproic acid	X	X	
verapamil hydrochloride	X	X	

4. Discussion

Known strengths and/or limitations of the method

The VCBA model requires different types of data, ranging from the biochemical, and physicochemical processes that occur in biological systems. Furthermore cell population or certain physiological and pathological conditions (example contamination) can influence a chemical reaction within the well and thus should be explored how to include these effects into the model prediction. These types of data are not always available from one source, which may lead to difficulties in establishing a reliable source of accurate and consistent information. Information gaps or poorly characterized processes can lead to unreliable simulations. Therefore, it is important to emphasize that the reliability of the simulations depends on the corresponding model and the uncertainty related to the incorporated data.

5. Contact Details

Dr. Andrew Worth

Chemicals Safety and Alternative Methods Unit - European Union Reference Laboratory for Alternatives to Animal Testing (EURL-ECVAM)
European Commission - Joint Research Centre
Corresponding author
via E. Fermi 1
Ispra (VA) 21027
email: andrew.worth@ec.europa.eu
telephone: +390332783527

Dr. Alicia Paini

Chemicals Safety and Alternative Methods Unit - European Union Reference Laboratory for Alternatives to Animal Testing (EURL-ECVAM)
European Commission - Joint Research Centre
Method developer / user
via E. Fermi 1
Ispra (VA) 21027
telephone: 39 0332 7813527

Dr. Jose Vicente Sala Benito

Chemicals Safety and Alternative Methods Unit - European Union Reference Laboratory for Alternatives to Animal Testing (EURL-ECVAM)
European Commission - Joint Research Centre
Corresponding author, Method developer / user
via E. Fermi 1
Ispra (VA) 21027

telephone: 39 0332 7813527

Ver n°	Date	Name	Type of change
1	10 February 2015	Dr. Alicia Paini	Method summary created

6. Abbreviations & Definitions

As	medium-headspace exchange surface area, m ² ,
Al	headspace losses surface area, m ² ,
alfa	a parameter linked to the in silico simulation of the mmp values.
C	Concentration, mg/m ³ ,
[C]	concentration of cell lipid in medium, kg m ⁻³ ,
C _{aq}	the concentration in the aqueous phase,
C _b	concentration in the cell, g.gw.w ⁻¹ ,
C _{mit}	the concentration inside of the mitochondria,
COSMOS	EU integrated project: COSMOS develops an integrated suite of computational tools to predict the effects of long-term exposure to chemicals in humans based on in silico calculations and estimation techniques. Website: http://www.cosmos-tox.eu/
D	diffusion coefficient, m ² s ⁻¹ ,
d _i	duration (in hours), within the i-th stage,
F	mass flow, mg m ⁻² s ⁻¹ ,
F	the Faraday constant
f _i	refer to the mass fraction of each compartment (aqueous, lipid, proteins) in the cell,
H	Henry law constant, Pa m ³ mol ⁻¹ ,
KC	cells-medium partitioning coefficient, m ³ kg ⁻¹ lipid,
KGL	gas-liquid distribution coefficient,
k _{mit}	diffusive parameter in mitochondria depending on the chemical and cell line applied,
KNIME	Konstanz Information Miner, is an open source data analytics, reporting and integration platform, which integrates various components for machine learning and data mining through its modular data pipelining concept. Website: www.knime.org
K _{ow}	octanol-water partition coefficient
K _P	plastic-medium partitioning coefficient, m,
K _S	partitioning coefficient between serum and medium, m ³ mol ⁻¹ ,
k _{AW}	two film mass transfer coefficient, m s ⁻¹ ,
k _G	air film mass transfer coefficient, m s ⁻¹ ,
k _L	liquid film mass transfer coefficient, m s ⁻¹ ,
k _t	killing rate,
k	reaction rate constant, s ⁻¹ ,
mmp	mitochondrial membrane potential
MW	molecular weight, g/mol
n	number of moles of a compound
NEC	no effect concentration
P	pressure, Pa,
QIVIVE	Quantitative In vitro to in vivo extrapolation
R	universal gas constant

R	A programming language and software environment for statistical computing and visualisation. Website: http://www.r-project.org/
Sc	Schmidt number,
[S]	concentration of proteins in medium, mol m ⁻³ ,
SEURAT	EU integrated projects cluster. SEURAT-1 Research Initiative develops a long-term research strategy leading to pathway-based human safety assessments in the field of repeated dose systemic toxicity testing of chemicals. Website: http://www.seurat-1.eu/
SM	plastic- medium exchange surface area, m ²
t	time, s,
T	temperature, K
V	volume, m ³ or in liter L,
W	organism wet weight, g,
zi	mortality rate, h.

Superscripts used in the formula's

diss	dissolved
p	plastic
S	serum
T	total

Subscripts used in the formula's

AW	air-water
degr	degradation
I	loses
M	medium
H	headspace

7. Bibliography

Method development:

- Paini A., et al., (in preparation)
A new refined Virtual Cell Based Assay for simulation of in vitro intracellular concentration,
Toxicology In Vitro Special Issue
- Sala Benito JV. et al., (in preparation)
Automated workflows for in silico modelling of chemical fate and toxicity,
Toxicology In Vitro Special Issue
- Zaldivar Comenges JM. et al., (2016)
Theoretical and mathematical foundation of the Virtual Cell Based Assay - A review.
Toxicology In Vitro Special Issue, DOI: 10.1016/j.tiv.2016.07.013
Zaldivar, J.M., Menecozzi, M., Macko, P., Rodriguez, R., Bouhifd, M., Baraibar Fentanes, J., (2011)
A biology-based dynamic approach for the modelling of toxicity in cell assays. Part II: models for cell population growth and toxicity.
[JRC publication repository](#)
- Zaldivar, J.M., Menecozzi, M., Marcelino Rodrigues, R., Bouhifd, M., (2010)
A biology-based dynamic approach for the modelling of toxicity in cell-based assays. Part I: fate modelling.
[JRC publication repository](#)

General references:

- Alberts B, Johnson A, Lewis J, et al. (2002)
The Mitochondrion, Molecular Biology of the Cell. 4th edition. New York: Garland

- Kramer, N. I. (2010)
Measuring, modelling and increasing the free concentration of test chemicals in cell assays. *PhD Thesis*.
Utrecht University.

Last update: 10 February 2015

DB-ALM Method summary n° 163 : The Human Bio-Accumulation Model

Systemic Toxicity, Bioavailability

The **Human Bio-Accumulation Model** is a mathematical model for screening of chemicals for human bioaccumulative potential using a simple physiologically based toxicokinetic model, which ranks chemicals based on their human bioconcentration factor (hBCF).

1. General Information

Corresponding author(s):

Dr. Andrew Worth, European Commission - Joint Research Centre

Date of data sheet creation: 10 February 2015

Last version number: 1

Last update: 10 February 2015

Section **No.5 Contact Persons** provides a complete list of the authors and method developers.

The method summary presented here describes a mathematical model.

In the context of this document, the term "method" in the paragraph headings refers to the hBCF PBK model.

Proprietary and/or Confidentiality Issues:

None. Work carried out at the Joint Research Centre with partial support from the FP7 COSMOS project.

Status

Development completed

The hBCF - PBK model for simulation of single and repeat exposure is built in R and translated in KNIME (the Konstanz Information Miner) automated workflows, as a part of the COSMOS project (Integrated *In Silico* Models for the Prediction of Human Repeated Dose Toxicity of Cosmetics to Optimise Safety), within the SEURAT-1 Research Initiative (Safety Evaluation Ultimately Replacing Animal Testing 1).

More information on the COSMOS project is available on its dedicated website:

<http://www.cosmostox.eu>



The KNIME workflows and documentation are available via COSMOS Space:

<http://cosmosspace.cosmostox.eu>

A web version of the models was also implemented, allowing to execute the models via a web browser. It is available through the COSMOS KNIME WebPortal:

<http://knimewebportal.cosmostox.eu>

2. Method Definition

Purpose of the Method

TYPE OF TESTING	:	Screening, Part of an integrated testing strategy
LEVEL OF TOXICITY ASSESSMENT	:	Risk assessment
PURPOSE OF TESTING	:	Safety

Context of use

Toxicologists evaluating personal care products and cosmetics have the possibilities to find within the product several harmful ingredients such as pesticides, carcinogens, reproductive toxicants, endocrine disruptors, plasticizers, degreasers and surfactants. To prevent cases in which the presence of a contaminant can result in unacceptable consequences for human health, it is generally agreed that risk analysis has to be carried out on the specific chemical. However, this is not always feasible due to the large number of existing chemicals and, for many of them; there is a lack of detailed information on their behavior in the environment and to human exposure. Generally screening approaches are necessary to prioritize and to select chemicals of concern for a more detailed risk characterization and analysis (Muir and Howard 2006; Daginnus et al., 2011).

Substances identified during a screening phase as PBT or vPvB - (very) persistent, (very) bioaccumulative are of the highest concern and subjected to a detailed risk assessment (ECHA, 2014).

The risk assessment paradigm is based on two aspects: first, the hazard of a given chemical, which ideally can be related to its physico-chemical properties, and second its exposure levels to ecosystems and to humans. A comparison of exposure levels and safe chemical doses/concentrations is then carried out and the risk is evaluated. Hazards of a substance are evaluated according to its persistence in the environment, long-range transport potential (LRTP), bioaccumulation characteristics and toxicity (Klasmeier et al. 2006). An important characteristic which was addressed in the work of Tonnelier et al., (2012) was to assess the screening of chemicals for human bioaccumulative potential using a simple physiologically based toxicokinetic model, which ranks chemicals based on their human bio concentration factor (hBCF).

A generic physiologically based kinetic (PBK) model, based on *in vitro* human liver metabolism data, minimal renal excretion and a constant exposure, was developed to predict the bioaccumulation potential of a chemical. The aim of the presented method is to provide an open source approach for assessing the potential of chemicals to bioaccumulate in humans following realistic exposure scenarios.

Target Users and Uses

The model can be used by Risk Assessor, when performing risk assessment, to rank potency of chemicals based on their BCF, and thus give priority for testing (refinement of *in vitro* strategy testing), especially if found in traces as contaminants in cosmetics products.

Scientific Principle of the Method

Model components

The hBCF - PBK model has been developed by the JRC, based on the initial and original formulation of Tonnelier et al. (2012). The original model has been refined, extended, and recoded for implementation in the KNIME platform.

As a starting point a generic PBK model was used that, based on *in vitro* liver metabolism data, minimal renal excretion and a chronic exposure, is able to assess the bioaccumulative potential of a chemical. The approach has been analysed using literature data for some well-known bioaccumulative compounds, data from the ECVAM (European Centre for the Validation of Alternative Methods) database, and for a subset of the ToxCast phase I chemical library.

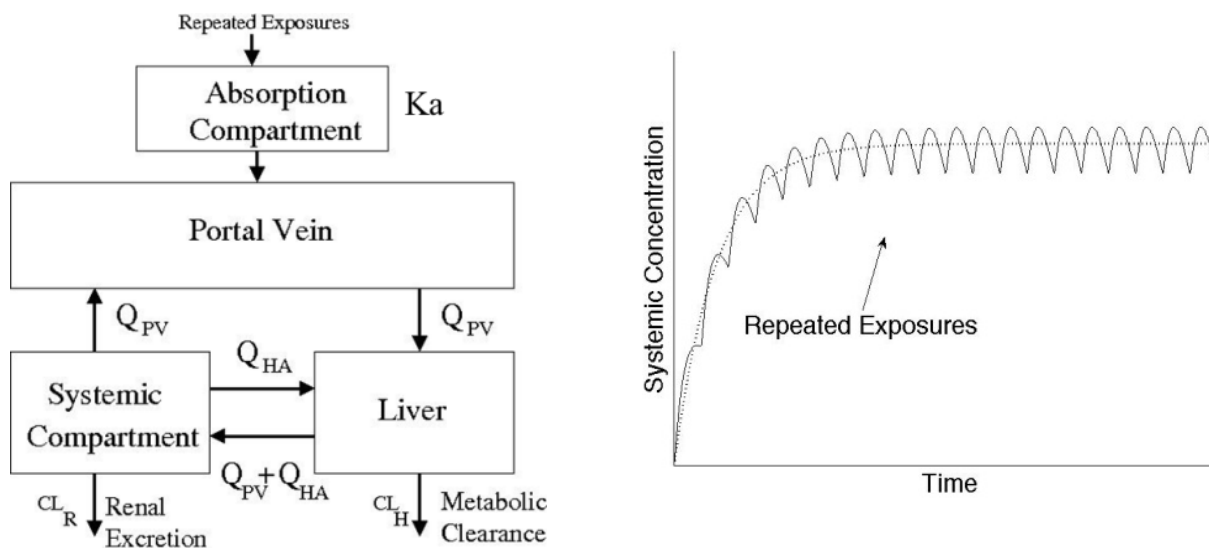


Figure 1. *Left:* Schematic representation of a minimal PBK model used to simulate perfusion-limited uptake of compound under a chronic exposure; *Right:* time profile of the increase in blood concentration of the chemical with repeat exposures. The dotted line is the trajectory of the averaged PBK model (from Tonnelier et al., 2012).

Procedure Description

Biological Endpoint and Measurement:

Bioavailability

Systemic availability

Bioaccumulation - the continuous increase in the concentration of a chemical in an organism, compared to the chemical's concentration in the environmental media to which the organism is exposed, i.e., air, water, soil, food, etc.

Endpoint Value:

Human Bioconcentration Factor (hBCF)

Basic procedure:

The hBCF PBK model is a mathematical approach. The model is coded in R and implemented in the open-source KNIME platform. The workflow is available through the COSMOS space:

- <http://cosmosspace.cosmostox.eu/app/space/>

The equations used in the individual model components are described in the next section: **2.1 Model description**.

A presentation on the COSMOS Automated *in silico* tools for *in vitro* to *in vivo* extrapolation was recorded for the American Society for Cellular and Computational Toxicology (ASCCT). Method developer Dr. Alicia Paini explains the composition of the model and the use of the KNIME workflows:

- <https://youtu.be/vJc1UEbuWdE>

2.1 Model Description

Model Equations

Based on Tonnelier et al. (2012) the simplified PBK model was re-written in R. The three compartment physiologically based model equivalent to the minimal Simcyp model (Simcyp Limited, Sheffield, UK) was implemented. The absorption of the substance is assumed to occur at different times (t_i) at a small dose (noted D), and the elimination occurs through renal filtration and liver metabolism (Tonnelier et al., 2012). The toxicokinetics of each substance is described by a system of differential equations of the following form, and taking into account the uptake that appears in the portal vein compartment using a supplementary compartment, leading to the following coupled system of differential equations:

$$V_{PV} \cdot \frac{dC_{pv}}{dt} = Q_{PV} \cdot (C_{SYS} - C_{PV}) + \sum_{t_i} fa \cdot F_G \cdot ka \cdot D \exp(-ka(t-t_i)) H(t-t_i) \quad [1A]$$

$$V_{PV} \cdot \frac{dC_{pv}}{dt} = Q_{PV} \cdot (C_{SYS} - C_{PV}) + k_a \cdot A(t) \quad [1B]$$

$$V_{PV} \cdot \frac{dC_{PV}}{dt} = Q_{PV} \cdot (C_{SYS} - C_{PV}) + D/T \quad [1C]$$

$$\begin{aligned}
 V_{liv} \cdot \frac{dC_{liv}}{dt} &= Q_{PV} \cdot (C_{PV} - C_{liv}) + Q_{HA} \cdot (C_{SYS} - C_{liv}) - CL_H \cdot C_{liv} \\
 V_{SS} \cdot \frac{dC_{SY}}{dt} &= Q_{PV} \cdot (C_{liv} - C_{SYS}) + Q_{HA} \cdot (C_{liv} - C_{SYS}) - CL_R \cdot C_{SYS}
 \end{aligned}
 \tag{2}$$

where

- C_{PV} , C_{liv} and C_{SYS} are the concentrations of the compound in the portal vein, liver and systemic circulation, respectively;
- V_{PV} and V_{liv} are the volumes of portal vein blood and liver, respectively,
-
- Q_{PV} and Q_{HA} are the blood flows in the portal vein and hepatic artery;
- fa is the fraction of the compound absorbed into the enterocytes;
- F_G is the intestinal availability of the substrate;
- k_a is the first-order absorption rate constant;
- H is the Heaviside step function;
- CL_H is the hepatic clearance and
- CL_R is the renal clearance.

Model developers considered here a complete absorption, $fa = 1$, and a full intestinal availability $F_G = 1$. The rate of renal excretion is estimated by multiplying the percent of unbound chemical by the normal adult glomerular filtration rate ($Q_f = 6.7 \text{ l/h}$) (Rotroff et al., 2010). The hepatic clearance CL_H is based on the whole-blood toxicant concentration and is derived from the experimentally measured intrinsic metabolic clearance ($CL_{u_{int}}$), based on unbound drug concentration.

The route of exposure does not appear explicitly in the averaged model and the amount of compound, D , could be seen as the result of the sum of different uptake routes (oral and dermal for instance). After a transient, the averaged model stabilizes on a steady state (C_{PV}^* , C_{liv}^* , C_{SYS}^*) from which the human bioconcentration factor is calculated as:

$$hBCF = \frac{C_{SYS}^*}{D/T} V_{PV}/t \tag{3}$$

The concentration in the systemic compartment is used to measure the extent to which a compound accumulates in the body and is compared with a theoretical concentration of compound in the portal vein, assuming a complete absorption of the dose D . The term D/V_{PV} reflects the concentration of the compound in the portal vein where the volume of portal vein is used as a reference. This choice stems from the occurrence of the dose D in the kinetics equation of the portal vein where the dose D occurs through the rate flow D/T . The quantity t in Equation [3] is a time set to one hour. Therefore, the factor V_{PV}/t is a normalizing factor that provides a dimensionless bioaccumulation factor.

The steady-state is reached when the time-variations of the concentrations vanish and it is straightforward to calculate C_{SYS} from [2] by equating to zero all the rate of change in the different compartments (left-hand side of system [2]). From [3], the approximated value of $hBCF$ is found:

$$hBCF = \frac{V_{PV}}{CL_H + fuQ_f \left(1 + \frac{CL_H}{Q_{PV} + Q_{HA}} \right)} \tag{4}$$

where $CL_R = fuQ_f$ is used. Neither the dose D nor the duration T occur in the analytical expression and therefore the proposed metric for bioaccumulation assessment (Expression [4]) is, on average, independent of the exposure scenario used to compute it.

Input data: Model Parameters

Chemical Parameters: MW, LogKow, pKa.

Additionally the following parameters are needed:

- The blood ratio (which value is set for all chemical to 0.55),
- Cl_{int} - the clearance values for liver ($\mu\text{M}/\text{min}/10^6$ cells) and kidney (L/h),
- F_u - unbound fraction of chemical in plasma,
- F_{mic} - the free fraction of compound in the microsomal preparation.
- C_{sys}^* steady-state concentration
- T_{sys} time in h,

PBK model Parameters:

Q_{PV} = portal vein flow, 63 L/h.
 Q_{HA} = hepatic arterial flow, 18 L/h.
 Q_F = Standard glomerular filtration rate, 6.7 L/h.
 V_{PV} = Volume portal vein, 3.2 L.
 V_{liv} = Volume Liver, 1.75 L.
 V_{sys} = Volume systemic circulation, 1 L.
 F_G = intestinal availability (absorption rate), 1.
 fa = absorption rate, 1.
Liver weight = liver weight, 1500 g
 D = dose, 0.01 mg.
 T = dose interval, 0.1 h.
Init = initial time point (simulation), 0 h.
End = end time point (simulation), 5 h.

3. Method Performance

Robustness of the Method

General performance measures:

The hBCF PBTK model underwent an extensive evaluation, described in the next section: **3.1 Evaluation of Model Performance.**

PBK Model Validation

The PBK model described in Tonnelier et al. (2012) was validated as follows: the PBK model simulations for twelve chemicals were compared with published pharmacokinetic/toxicokinetic data and published PBK model predictions.

The results agree with those obtained by Rotroff et al. (2010) concerning 2,4-dichlorophenoxyacetic acid, oxytetracycline dehydrate, triclosan, bisphenol A and parathion.

Simulated results, at 40 h, agree with experimental data on plasma elimination half-lives of warfarin for the two enantiomers: R-warfarin 46 ± 7 h and S-warfarin 36 ± 13 h, respectively.

The prediction of the maximum plasma concentration for chlorpyrifos and propranolol hydrochloride was similar to human kinetic studies and published PBPK model results. In the last case, also the data on the time at which C_{max} was reached are gathered and compared with the simulation, i.e., 2 h in the two experiments (10 and 20 mg day⁻¹ oral dose) and 2.1 h in the simulations.

Apart from the well-known persistent compounds such as PCBs, DDT and PFOS were uncertainty on the estimations (Simcyp could only run for 200 days, and the values were extrapolated), and the experimental variability (e.g. for PCB-153, published values between 5 and 27.5 years were found) tends to be higher, the main discrepancy is obtained for thioridazine where the predicted elimination half-life

value is much higher than the experimental one: an elimination half-life that oscillates around 26 h has been reported while the model predicted 87 days.

This points out a limitation of the approach due to the fact that the model only considers liver metabolism and minimal renal excretion, whereas the main excretion route of thioridazine seems to be through the faeces.

A similar limitation holds for oxytetracycline where the observed overprediction of the concentration possibly lies in the fact that a 100% oral absorption has been considered, whereas a low oral bioavailability has been reported for this compound (Nielsen and Gyrd-Hansen 1996). However, as stated before, the main interest of model developers was a fast screening procedure to estimate human bioaccumulation potential for risk assessment and, as a conservative approach, false-positive predictions were not main concern.

Test Compounds

In Tonnelier et al. (2012), the chemicals selected were: 55 organic chemicals, mostly drugs and pesticides, which is a subset selected from the list of an international ICCVAM validation (2009), and a subset of 35 substances (Rotroff et al. 2010), mostly pesticides, of the ToxCast Phase I chemical library (<http://www.epa.gov/ncct/toxcast/chemicals.html>). For the ECVAM database chemicals, liver metabolism and unbounded fraction data were taken from Pelkonen et al. (2009) and Rousu et al. (2010), whereas a similar data for the chemicals in ToxCast phase I have been published in Rotroff et al. (2010). The list of selected chemicals as well as their physico-chemical parameters can be found as the Supplementary Material in Tonnelier et al. (2012).

4. Discussion

Known strengths and/or limitations of the method

Due to the high number of substances produced by the chemical industry, there exists the need to prioritise them to focus the risk analysis effort. Therefore, the aim of a chemical prioritization exercise is to screen a considerable number of substances to select the ones for which a detailed risk assessment would be necessary according with the type of selection criteria defined beforehand.

Bioaccumulative potential (B) is an important element in all exercises of chemical prioritization. The approach followed (if no experimental data are available) is based on the calculation of the lipophilicity of the substance, sometimes based on correlations between a certain human Bioconcentration Factor (hBCF) and the octanol-water partition coefficient (Kow). Even though this approach takes into account the fact that hydrophobic compounds tend to bioaccumulate in lipids, it does not consider the processes that will tend to decrease the concentration of the compound in a certain organisms, such as excretion, depuration and/or metabolization processes.

A direct and straightforward estimation of the hBCF based solely on a limited number of compound parameters would be of advantage in prioritization of chemicals and may provide an efficient pre-screening criterion for a rapid assessment, this approach can be applied in the assessment of cosmetics ingredients. Tonnelier et al., (2012) developed a simplified mapping based on *in vitro* human liver metabolism data, minimal renal excretion and a constant exposure. Two parameters to assess the bioaccumulation potential of compounds: clearance and protein-binding affinity were measured *in vitro* to be used as model input. The proposed measure of hBCF is derived from the steady state reached by the PBK model after successive exposures.

This model was designed to incorporate not only the chemical properties of the compounds, but also the processes that tend to decrease the concentration of the compound, such as metabolism. Following this work the generic PBK model was implemented and re-written in R, in the open source KNIME interface using MySQL database. The outputs are:

1. a graphical figure expressed as concentration versus time;
2. the systemic concentration (C_{sys}) at steady state of the selected compound; and
3. the human bioconcentration factor (hBCF).

The ability to describe the bioaccumulative potential of a compound through the metabolism combined with the renal excretion shows that the clearance is the main determinant factor for human bioaccumulation.

The vast majority of the approaches to predict the bioaccumulation potential consider only the partitioning of the chemical and not the biotransformation potential of the organism. The results obtained in this study suggest that it is possible to have a pre-assessment of the human potential bioaccumulation by performing two *in vitro* tests: one to calculate the chemical binding to plasma proteins and the other to estimate the liver clearance by *in vitro* measuring the metabolite formation and/or the substrate depletion of the compound using human hepatocytes.

Both types of tests are suited to high-throughput analysis and therefore can be used for screening purposes for the prioritisation of new compounds. However, there are several aspects that should be considered carefully in this approach. First, it is assumed that the *in vitro* measurements adequately describe the *in vivo* activity. Second, the experimental conditions in the *in vitro* tests should have been properly selected, e.g., compounds concentration low enough to avoid saturation when measuring substrate depletion. Third, concerning the *in vitro* clearance, it is assumed that there is no further biotransformation in the gastrointestinal tract and, in addition, metabolism occurs only in the liver.

Normally, all these assumptions are conservative, and therefore, it is more probably to obtain false positives than false negatives as it can be observed in the compounds that deviate in the comparison between experimental data and predictions.

Potential for future development

Risk assessors could use this simple PBK model to rank chemicals based on their potency, and thus give priority for testing (refinement of *in vitro* strategy testing), especially if found in traces as contaminants in cosmetics products. Furthermore, there is the need to develop sensitive throughput analytical techniques that can allow a better understanding of the kinetics aspects in *in vitro* experiments. This would also help in determining a general procedure for performing *in vitro*–*in vivo* extrapolation (IVIVE) and being able to move from concentration to dose response without the use of animal experiments (Adler et al. 2011). This IVIVE approach could be applied to cosmetics ingredients and would be based on their potential to bio-accumulate in human.

The relevant dose would be calculated back by reverse dosimetry, using PBK modelling. This would be a good approach to be implemented in the cosmetics world; however, there is a lack of experimental data. These data could be easily produced using HTS techniques in a fast and reliable way.

Model developers recommended this approach, by first screening literature for available data for cosmetics ingredients and/or to generate new experimental data for the specific cosmetics ingredients, especially protein binding affinity and the hepatic clearance. It is therefore essential that *in vitro* metabolic clearance tests with metabolic competent cell lines be performed for suspected B compounds.

Conclusions and recommendations

- Dermal exposure is an essential exposure pathway and should be included in the present PBK model. This refinement of the model is at the moment being carried out at JRC, once the model is validated it could be used for prediction and potency ranking using the hBCF.
- To include Monte Carlo Simulation to take into account variability within the human population. To be able to use the model and scale it up to all human population.
- Develop more refined QSAR models to overcome uncertainties.

5. Contact Details

Dr. Andrew Worth

Chemicals Safety and Alternative Methods Unit - European Union Reference Laboratory for Alternatives to Animal Testing (EURL-ECVAM)
European Commission - Joint Research Centre
Corresponding author
via E. Fermi 1
Ispra (VA) 21027
email: andrew.worth@ec.europa.eu
telephone: +390332783527

Dr. Alicia Paini

Chemicals Safety and Alternative Methods Unit - European Union Reference Laboratory for Alternatives

to Animal Testing (EURL-ECVAM)
 European Commission - Joint Research Centre
 Method developer / user
 via E. Fermi 1
 Ispra (VA) 21027
 telephone: 39 0332 7813527

Dr. Jose Vicente Sala Benito

Chemicals Safety and Alternative Methods Unit - European Union Reference Laboratory for Alternatives to Animal Testing (EURL-ECVAM)
 European Commission - Joint Research Centre
 Method developer / user
 via E. Fermi 1
 Ispra (VA) 21027
 telephone: 39 0332 7813527

Ver n°	Date	Name	Type of change
1	10 February 2015	Dr. Alicia Paini	Method summary created

6. Abbreviations & Definitions

C_{PV}	the concentrations of the compound in the portal vein,
C_{liv}	the concentrations of the compound in the liver
C_{SYS}	the concentrations of the compound in the systemic circulation
CL_H	is the hepatic clearance and
CL_R	is the renal clearance.
$CL_{U_{int}}$	experimentally measured intrinsic metabolic clearance
COSMOS	EU integrated project: COSMOS develops an integrated suite of computational tools to predict the effects of long-term exposure to chemicals in humans based on in silico calculations and estimation techniques. Website: http://www.cosmos-tox.eu/
D	amount of compound,
ECVAM	European Centre for the Validation of Alternative Methods. As of 2011, ECVAM has been established as the European Union Reference Laboratory for Alternatives to Animal Testing (EURL ECVAM).
fa	the fraction of the compound absorbed into the enterocytes;
F_G	the intestinal availability of the substrate;
H	the Heaviside step function;
hBCF	human Bio-Concentration Factor
ICCVAM	Interagency Coordinating Committee on the Validation of Alternative Methods, a permanent committee of the NIEHS under the National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM). ICCVAM is composed of representatives from 15 U.S. Federal regulatory and research agencies that require, use, generate or disseminate toxicological and safety testing information.
ka	the first-order absorption rate constant;
KNIME	Konstanz Information Miner, is an open source data analytics, reporting and integration platform, which integrates various components for machine learning and data mining through its modular data pipelining concept. Website: www.knime.org
PBK	Physiologically Based Kinetic models
Q_{PV} and Q_{HA}	are the blood flows in the portal vein and hepatic artery;
QSAR	Quantitative Structure Activity Models

- R A programming language and software environment for statistical computing and visualisation. Website: <http://www.r-project.org/>
- SEURAT 1 EU integrated projects cluster. SEURAT-1 Research Initiative develops a long-term research strategy leading to pathway-based human safety assessments in the field of repeated dose systemic toxicity testing of chemicals. Website: <http://www.seurat-1.eu/>
- V_{PV} and V_{liv} are the volumes of portal vein blood and liver, respectively,
- V_{ss} is the volume of distribution of substrate;

7. Bibliography

Method development:

- Rotroff DM, Wetmore BA, Dix DJ, Ferguson SS, Clewell HJ, Houck KA, et al. (2010) Incorporating human dosimetry and exposure into high-throughput In vitro toxicity screening. *Toxicological Sciences* 117, 348-358
- Tonnelier A, Coecke S., Zaldivar JM. (2012) Screening of chemicals for human bioaccumulative potential with a physiologically based toxicokinetic model. *Archives of Toxicology* 86(3), 393-403
- Zaldivar JM, Marinov D, Dueri S, Castro-Jiménez J, Micheletti C, Worth A. (2011) An integrated approach for bioaccumulation assessment in mussels: towards the development of Environmental Quality Standards for biota. *Ecotoxicology and Environmental Safety* 74, 244-252

General references:

- Adler, S.; Basketter, D.; Creton, S.; Pelkonen, O.; van Benthem, J.; Zuang, V.; Andersen, K. E.; Angers-Loustau, A.; Aptula, A.; Bal-Price, A.; Benfenati, E.; Bernauer, U.; Bessems, J.; Bois, F. Y.; Boobis, A.; Brandon, E.; Bremer, S.; Broschard, T.; Casati, S.; Coecke, S.; Corvi, R.; Cronin, M.; Daston, G.; Dekant, W.; Felter, S.; Grignard, E.; Gundert-Remy, U.; Heinonen, T.; Kimber, I.; Kleijnans, J.; Komulainen, H.; Kreiling, R.; Kreysa, J.; Batista Leite, S.; Loizou, G.; Maxwell, G.; Mazzatorta, P.; Munn, S.; Pfuhler, S.; Phrakonkham, P.; Piersma, A.; Poth, A.; Prieto, P.; Repetto, G.; Rogiers, V.; Schoeters, G.; Schwarz, M.; Serafimova, R.; Tahti, R.; Testai, E.; van Delft, J.; van Loveren, H.; Vinken, M.; Worth, A. and Zaldivar, J.M. (2011) Alternative (non-animal) methods for cosmetics testing: current status and future prospects – 2010. [Link to final report](#) (last access 19.06.2013) [Link to key policy documents](#) (last access 19.06.2013) *Archives of Toxicology* 85, 367-485
- Daginnus K, Gottardo S, Payá-Pérez A, Whitehouse P, Wilkinson H, Zaldivar JM. (2011) A model-based prioritization exercise for the European Water Framework Directive. *8*, 435-455
- ECHA (2014) Guidance on Information Requirements and Chemical Safety Assessment Chapter R.11: PBT/vPvB assessment. [Link to document](#) (last access 07.05.2015) *European Chemicals Agency*
- ICCVAM (2009) Report on the ICCVAM-NICEATM/ECVAM/JaCVAM Scientific workshop on acute chemical safety testing: Advancing in vitro approaches and humane endpoints for systemic toxicity evaluations. Available from: http://iccvam.niehs.nih.gov/methods/acutetox/Tox_workshop.htm
- Klasmeier J, Matthies M, MacLeod M, Fenner K, Scheringer M, Stroebe M, et al. (2006) Application of multimedia models for screening assessment of long-range transport potential and overall persistence. *Environmental Science & Technology* 40, 53-60
- Muir DCG; Howard PH. (2006) Are there other persistent organic pollutants? A challenge for environmental chemists. *Environmental Science & Technology* 40, 7157-7166

- Pelkonen, O., Tolonen, A., Rousu, T., Tursas, L., Turpeinen, M., Hokkanen, J., Uusitalo, J., Bouvier d'Yvoire, M., and Coecke, S. (2009)
Comparison of metabolic stability and metabolite identification of 55 ECVAM/ICCVAM validation compounds between human and rat liver homogenates and microsomes - a preliminary analysis.
Alternatives to Animal Experimentation (ALTEX) 26, 214-222
- Rousu T, Hokkanen J, Pelkonen O, Tolonen A. (2010)
Applicability of generic assays based on liquid chromatography-electrospray mass spectrometry to study in vitro metabolism of 55 structurally diverse compounds.
Frontiers in Bioscience 1, 1-13

Last update: 10 February 2015

DB-ALM Method summary n° 168 : PPAR full agonists' pharmacophore-based predictor

Systemic Toxicity

The **PPAR full agonists' pharmacophore-based predictor** is an *in silico* model to predict the potential of chemicals to bind and activate PPAR γ (PPAR γ full agonism prediction). The model is based on a two-step procedure: docking of the chemical structures into the PPAR γ binding site and filtering using a pharmacophore model of PPAR γ full agonists.

1. General Information

Corresponding author(s):

Dr Ivanka Tsakovska, Institute of Biophysics and Biomedical Engineering, Bulgarian Academy of Sciences (BAS) Bulgaria

Date of data sheet creation: 16 November 2015

Last version number: 1

Last update: 16 November 2015

Section **No.5 Contact Details** provides a complete list of the authors and method developers.

The method summary presented here describes a mathematical model.

In the context of this document, the term "method" in the paragraph headings refers to the model.

Proprietary and/or Confidentiality Issues:

Public

Status

Development

Development completed. The PPAR γ full agonists' pharmacophore-based predictor was developed as a part of the COSMOS (Integrated In Silico Models for the Prediction of Human Repeated Dose Toxicity of Cosmetics to Optimise Safety), within the SEURAT-1 (Safety Evaluation Ultimately Replacing Animal Testing) Research Initiative. More information on the project is available on its dedicated website:

<http://www.cosmostox.eu>



2. Method Definition

Purpose of the Method

TYPE OF TESTING	:	Screening
LEVEL OF TOXICITY ASSESSMENT	:	Toxic potential

Context of use

PPAR γ is one of the important nuclear hormone receptors that contribute to excessive accumulation of triglycerides in hepatocytes. A prosteatotic AOP, integrating the ligand-induced activation of the PPAR γ as a molecular initiating event (MIE) has been proposed (Al Sharif et al., 2014). The PPAR γ full agonists' pharmacophore-based predictor is intended for prediction of this particular MIE. Thus it could be used for the *in silico* screening of full agonists of hepatic PPAR γ that can function as steatosis inducers.

Scientific Principle of the Method

Taking into consideration that ligand-dependent activation of PPAR γ can be a molecular initiating event (MIE) that triggers downstream events leading to liver steatosis, the molecular docking model can be used for *in silico* prediction of potential prosteatotic compounds that are PPAR γ full agonists. The *in silico* model was developed using data for PPAR γ full agonists extracted from Protein Data Bank (PDB) (3D receptor-ligand complexes, www.rcsb.org) and from scientific literature. It includes development of a pharmacophore model of PPAR γ full agonists and implementation of virtual screening (VS) protocol to dock the chemical structures into the PPAR γ binding site and to filter the developed pharmacophore.

Model Description

Structural and experimental data for 439 PPAR γ ligands were collected from PDB and the scientific literature (publicly available at <http://biomed.bas.bg/qsarmm/>). A subset of 170 PPAR γ full agonists was extracted from the initial dataset. The selected ligands constituted a structurally diverse dataset of PPAR γ full agonists with relative efficacy $\geq 70\%$ and/or PDB ligands with structural features fitting the developed PPAR γ full agonists' pharmacophore.

The pharmacophore model was built based on the three most active agonists extracted from PPAR γ PDB complexes (Tsakovska et al., 2014). It outlines seven important pharmacophore features that were observed in the most active agonists: four polar atoms and functional groups capable of performing hydrogen bonding and ionic interactions and three hydrophobic and aromatic structural elements. The seven-feature pharmacophore model was evaluated further among 21 full agonists from PDB, selected based on reliable data about their full agonistic activity. The analysis was performed by superposition of the agonist structures on the pharmacophore model and estimation of the correspondence between pharmacophore points and their respective substructures. Most of the compounds comprise 4–5 pharmacophore features. Thus for the purposes of the subsequent VS protocol the five-points pharmacophore model was used.

The VS protocol developed consists of three steps:

- i. protein preparation
- ii. docking of the ligands into the PPAR γ binding site
- iii. filtering of the generated poses based on the five-points pharmacophore model of PPAR γ full agonists.

The last step means the final poses that did not satisfy the pharmacophore were eliminated.

The pharmacophore was developed using the "Pharmacophore Query Editor" tool in molecular modelling (MM) platform MOE v. 2014.0901 (CCG Inc.). The application refers to the generation of a pharmacophore hypothesis for the binding interactions in a particular active site. In MOE, the computerised representation of a hypothesised pharmacophore is called a pharmacophore query. A MOE pharmacophore query is a set of query features that are typically created from ligand annotation points.

The receptor's ligand binding domain used in the docking protocol (X-ray structure of PPAR γ with rosiglitazone, PDB ID 1FM6) was initially prepared using the Protonate3D application in MOE. This application assigns the hydrogens following the optimal free energy proton geometry and ionisation states of titratable protein groups using the Generalized Born electrostatics model. The physiologically relevant parameters were set during the minimisation: temperature 310°K; pH = 7.4; ion concentration: 0.152 mol/L. The ligands were docked into the binding site of the prepared protein structure. The London dG scoring function without subsequent refinement was applied to score the poses of the docked ligands (MOE). The highly scored poses of each ligand with a negative value of the scoring function only were kept. The pharmacophore model was used as a placement method in the docking procedure.

Procedure Description

Biological Endpoint and Measurement:

Receptor binding: Peroxisome proliferator-activated receptor gamma 1 (PPAR γ 1)

Receptor binding: Peroxisome proliferator-activated receptor gamma 2 (PPAR γ 2)

Experimental System:

In silico mathematical model to predict the potential of chemicals to bind and activate PPAR γ 1 and PPAR γ 2, jointly referred to as PPAR γ

3. Method Performance

Test Compounds

Curated data from PDB and scientific literature consisted of 170 PPAR γ full agonists and 87 partial agonists that have been extracted from the initial dataset collected (structural and biological data for 439 PPAR γ ligands, publicly available at <http://biomed.bas.bg/qsar/mm/>). The threshold proposed by Henke et al. (1998) was adopted who considered full agonists those compounds that elicited in average at least 70% activation of PPAR γ as compared to rosiglitazone.

Predictive capacity:

The model was evaluated on different datasets:

- Application of the model on the set of 170 PPAR γ full agonists. In total 144 out of the 170 ligands were correctly predicted as full agonists (model sensitivity of 85%).
- Application of the model on a subset of 87 PPAR γ partial agonists. In total 38 did not pass the filter and were correctly classified as not being full agonists (model specificity in relation to the partial agonists of 44%).
- Application of the model on a subset of 2527 decoys randomly extracted from the full set of 25867 PPAR γ decoys (each 10th structure was selected after removal of duplicates) in DUD-E database (Directory of Useful Decoys - Enhanced, <http://dude.docking.org/>); of which 1949 were correctly classified as not being full agonists (model specificity in relation to decoys of 77%).

Applicability:

The model is based on publicly available structural and biological data for PPAR γ full agonists. Therefore it is applicable to structures possessing the typical pharmacophoric features of PPAR γ full agonists.

4. Discussion

Known experimental strengths and/or limitations of the method

The model is based on reliable 3D structural data and is applicable to broad range of chemical structures possessing the typical pharmacophoric features of PPAR γ full agonists.

The model is of broader toxicological significance since the MIE has been implicated in a range of adverse effects, including developmental and reproductive toxicity.

More conservative approach is applied meaning implementation of the less restrictive five-points pharmacophore model. This gives higher probability to predict false positives (partial agonists predicted as full) than false negatives.

In general scoring is the critical step in the docking simulations. There is no universal scoring function that is suitable for all types of binding pockets and ligands. Thus the prediction ability of the developed model is restricted by inherent limitation of the scoring function used in the docking procedure.

The application of the model requires availability of specialised molecular modelling software.

Technical requirements

Computer system capable of running MOE molecular modelling software.

Potential for future development

Further refinement of the model to differentiate better between full and partial agonists; elaboration of free docking software.

Comparisons to other methods

The present model is built on extended data set of full agonists covering most of the publicly available data in PDB and scientific literature. Thus it is applicable to compounds with much broader structural range.

5. Contact Details

Dr Ivanka Tsakovska

Department of QSAR & Molecular Modelling
Institute of Biophysics and Biomedical Engineering, Bulgarian Academy of Sciences (BAS)
Corresponding author, Method developer / user
105 G. Bontchev Str
Sofia 1113
Bulgaria
email: ITsakovska@biomed.bas.bg
telephone: +359 2 9793605

Dr Merilin Al Sharif

Department of QSAR & Molecular Modelling
Institute of Biophysics and Biomedical Engineering, Bulgarian Academy of Sciences (BAS)
Method developer / user
105 G. Bontchev Str
Sofia 1113
Bulgaria
email: merilin.al@biomed.bas.bg
telephone: +359 2 9793647

Petko Alov

Department of QSAR & Molecular Modelling
Institute of Biophysics and Biomedical Engineering, Bulgarian Academy of Sciences (BAS)
Method developer / user
105 G. Bontchev Str
Sofia 1113
Bulgaria
email: petko.alov@biophys.bas.bg
telephone: +359 2 9793647

Dr Ilza Pajeva

Department of QSAR & Molecular Modelling
Institute of Biophysics and Biomedical Engineering, Bulgarian Academy of Sciences (BAS)
Method developer / user
105 G. Bontchev Str
Sofia 1113
Bulgaria
email: pajeva@biomed.bas.bg
telephone: +359 2 9793605

Ver n°	Date	Name	Type of change
1	16 November 2015	Dr Ivanka Tsakovska	Data sheet created

6. Abbreviations & Definitions

AOP:	Adverse Outcome Pathway
COSMOS:	EU Integrated Project. COSMOS developed an integrated suite of computational tools to predict the effects of long-term exposure to chemicals in humans based on <i>in silico</i> methods. Website: http://www.cosmostox.eu
Decoy:	Compounds that are selected to resemble the receptor binders' physicochemical properties but at the same time are topologically dissimilar to them in order to minimise the likelihood of the actual binding
Docking:	Modelling technique that positions the ligand in different orientations and conformations within the binding site to calculate optimal binding geometries and energies
DUD-E database:	Directory of Useful Decoys - Enhanced
MIE:	Molecular Initiating Event
MM:	Molecular Modelling
PDB:	Protein Data Bank http://www.rcsb.org/pdb/home/home.do
Pharmacophore:	An ensemble of steric and electronic features that is necessary to ensure the optimal interactions with a specific biological target and to trigger (or block) its biological response
PPAR γ :	Peroxisome Proliferator-Activated Receptor γ
SEURAT-1:	Safety Evaluation Ultimately Replacing Animal Testing 1
VS:	Virtual Screening

7. Bibliography

Method development:

- Al Sharif M., Tsakovska I., Pajeva I., Alov P., Fioravanzo E., Bassan A., Kovarich S., Yang C., Mostrag-Szlichtyng A., Vitcheva V., F, Worth A.P., Richarz A.-N., Cronin MTD. (2016)
The application of molecular modelling in the safety assessment of chemicals: A case study on ligand-dependent PPAR γ dysregulation.
Toxicology In press, <http://dx.doi.org/10.1016/j.tox.2016.01.009>
- Al Sharif, M., Alov, P., Vitcheva, V., Pajeva, I., Tsakovska, I. (2014)
Modes-of-Action Related to Repeated Dose Toxicity: Tissue-Specific Biological Roles of PPAR γ Ligand-Dependent Dysregulation in Nonalcoholic Fatty Liver Disease
PPAR Research 2014, 432-647
- COSMOS (2016)
<http://www.cosmostox.eu/>
- Fratev F., Tsakovska I., Al Sharif M., Mihaylova E., Pajeva I. (2015)
Structural and dynamical insight into PPAR γ antagonism: in silico study of the ligand-receptor interactions of non-covalent antagonists.
International Journal of Molecular Sciences 16(4), 15405-15424
- Henke, B.R., Blanchard, S.G., Brackeen, M.F., Brown, K.K., Cobb, J.E., Collins, J.L., Harrington, W.W., Hashim, M.A., Hull-Ryde, E.A., Kaldor, I., et al. (1998)
N-(2-Benzoylphenyl)-L-tyrosine PPAR γ agonists. 1. Discovery of a novel series of potent antihyperglycemic and antihyperlipidemic agents.
Journal of Medicinal Chemistry 41, 5020–5036.
- Molecular Operating Environment (MOE) (2015)
version 2014.091; Chemical Computing Group Inc., 2015, <http://www.chemcomp.com>.
- Protein Data Bank (PDB) (2016)
RCSB Protein Data Bank - www.rcsb.org (access 21.07.2016)

- Tsakovska, I., Al Sharif, M., Alov, P., Diukendjieva, A., Fioravanzo, E., Cronin, M., Pajeva, I. (2014)
Molecular Modelling Study of the PPAR γ Receptor in Relation to the Mode of Action/Adverse Outcome
Pathway Framework for Liver Steatosis.
International Journal of Molecular Sciences 15, 7651–7666

Last update: 16-11-2015

DB-ALM Method summary n° 169 : QSAR model for the prediction of LXR binding potential.

Systemic Toxicity

This quantitative structure-activity relationship ((Q)SAR) model is a classification model (based on PLS-DA) for the prediction of Liver X Receptor (LXR) binding potential. The binding and activation of LXR has been identified among the molecular initiating events leading to liver steatosis.

1. General Information

Corresponding author(s):

Dr Simona Kovarich, S-IN Soluzioni Informatiche S.r.l. Italy

Date of data sheet creation: 03 November 2015

Last version number: 1

Last update: 03 November 2015

Section **No.5 Contact Details** provides a complete list of the authors and method developers.

The method summary presented here describes a mathematical model.

In the context of this document, the term "method" in the paragraph headings refers to the model.

Proprietary and/or Confidentiality Issues:

Public

Status

Development

Development completed. Developed in the FP7 COSMOS Project, which was part of the Research Initiative SEURAT-1: Towards the Replacement of In Vivo Repeated Dose Systemic Toxicity Testing (2011-2015). More information on the COSMOS project is available on its dedicated website:

<http://www.cosmostox.eu>



The model is implemented as a KNIME workflow, available on the COSMOS project COSMOS Space website:

<http://cosmosspace.cosmostox.eu>

A web version of the model was also implemented, allowing to execute the model via a web browser. It is available through the COSMOS KNIME WebPortal:

<http://knimewebportal.cosmostox.eu>

A tutorial is available at:

<https://www.youtube.com/watch?v=EkB1i9a4TWI>

2. Method Definition

Purpose of the Method

TYPE OF TESTING	:	Part of an integrated testing strategy
LEVEL OF TOXICITY ASSESSMENT	:	Toxic potential
PURPOSE OF TESTING	:	Screen or priority setting

Context of use

The binding and activation to LXR has been identified among the molecular initiating events leading to liver steatosis (Landesmann et al., 2012). Thus, a QSAR model for the prediction of potential LXR binding may be used as part of integrated approaches to testing and assessment (IATA)/integrated testing strategies (ITS) to screen chemicals potentially sharing the same Adverse Outcome Pathway (AOP) (i.e. from LXR activation to liver steatosis). The model may also find application in the grouping of chemicals to support read-across justification, in terms of mechanistic plausibility (as molecular initiating event leading to long-term adverse effects in the liver), and weight of evidence evaluations.

The QSAR model has been developed according to the OECD principles for QSAR validation for regulatory purposes (OECD, 2006; OECD, 2007). The QMRF (QSAR Model Reporting Format) is made available in COSMOS Space (<http://cosmospace.cosmostox.eu>) and will be submitted for publication in the JRC (Q)SAR Model Database.

Scientific Principle of the Method

The binding and activation to LXR has been identified among the molecular initiating events involved in the adverse outcome pathway leading to liver steatosis.

A (quantitative) structure-activity relationship ((Q)SAR) model is a mathematical model relating one or more quantitative parameters derived from a chemical structure to a quantitative measure (continuous or categorical) of a property or activity. This QSAR model predicts potential LXR binding by classifying chemicals as potential binders ("active") or not-binders ("inactive"). The QSAR is based on a set of few molecular descriptors encoding for basic electronic properties, hydrophobicity, molecular shape and complexity. These features are known to be important for binding to the hydrophobic ligand binding pocket of LXR receptor.

Model Description

The prediction model provides a classification of potential binders and not-binders to LXR.

The dataset used for model development consisted of 97 chemicals selected from a wider dataset of 356 known LXR binders (biological data and structure collected from the literature and available database). The dataset included 50 ligands characterised by an $IC_{50} < 20$ nM, which were assigned to the ACTIVE class, and 47 ligands with an $IC_{50} > 1000$ nM, which were assigned to the INACTIVE class. A detailed description of dataset assembling and class assignment is provided in the model QMRF.

After a preliminary explorative analysis of the dataset (e.g., response distribution, physico-chemical and 2D structural space PCA analysis), several PLS-DA (Partial Least Square – Discriminant Analysis) classification models were developed using the commercial in-house software SIMCA (SIMCA ver. 13.0.3.0, 2012) based on different combination of MOSES molecular descriptors (MOSES.Descriptors, 2011 - www.molecular-networks.com), e.g. physico-chemical and 1D-2D-3D descriptors. The classification performance was evaluated using Cooper statistics (Cooper et al., 1979) and the following equations:

Sensitivity: $Sn = TP / (TP + FN)$

Specificity: $Sp = TN / (TN + FP)$

Overall Accuracy: $OA = (TP + TN) / (TP + FN + FP + TN)$

where: TP (true positive) and TN (true negative) are the number of compounds correctly classified as active and inactive, respectively; FN (false negative) is the number of active compounds classified as inactive, and FP (false positive) is the number of inactive compounds classified as active. As a precautionary principle, the misclassification of active compounds as inactive (false negative) should be considered a much greater error than the misclassification of inactive compounds as active (false positive). For this reason the number of FN was minimized for the selection of the best models.

Models were internally validated by means of 7-fold full cross-validation, and externally validated by means of an a priori 30% splitting of the dataset into training (78 compounds) and test set (19 compounds). One PLS-DA classification model was finally selected based on classification accuracy (optimizing sensitivity), external predictivity, model interpretability and reproducibility. The model was based on three latent variables (A) derived from seven MOSES 2D descriptors (selected based on the

parameter Variable Influence on Projection, that maximizes Q2CV), which encode basic electronic properties, hydrophobicity, molecular shape and complexity: HDon_O (Number of oxygen atom-based hydrogen bonding donors), Polariz (polarizability), NRotBond (Number of rotatable bonds), NAtoms (Number of atoms), NStereo (Number of tetrahedral stereo centers), Complexity (Molecular complexity), Rgyr (radius of gyration).

To allow for model applicability and reproducibility, the QSAR PLS-DA model was implemented into KNIME (Berthold et al., 2007) and a workflow was developed for model development, validation (cross-validation (CV) and external), prediction generation (also for new chemicals) and applicability domain definition. In particular, two approaches were used to define the applicability domain of the model: i) Leverage approach, ii) Similarity approach (based on Euclidean distances). Applicability domain thresholds (ADT) were calculated for each approach employing two Enalos KNIME nodes (<http://www.novamechanics.com/knime.php>), namely Enalos Domain-Leverage (Melagraki et al., 2009; Afantitis et al., 2008) and Enalos Domain-Similarity (Melagraki et al., 2010; Afantitis et al., 2011).

ADT for Leverage = 0.231

ADT for Similarity = 16.55

Compounds with leverage/similarity values greater than the specified thresholds are considered outside the applicability domain of the model and their predictions could be not reliable since model extrapolations.

To make the model freely available, two related PLS-DA models based on freely available molecular descriptors, namely PaDEL (Yap, 2011) and RDKit descriptors, were developed and implemented as KNIME workflows (Berthold et al., 2007). Six PaDEL descriptors and five RDKit descriptors, highly correlated to the seven MOSES modelling descriptors and encoding for similar structural information (e.g., hydrophobicity, molecular shape and complexity), were selected as modelling variables. The PLS-DA classification models implemented in KNIME showed similar classification performance as compared to the original QSAR model developed in SIMCA. In particular, the QSAR classification model based on PaDEL descriptors include the following six modelling descriptors: AMR (molar refractivity), nHeavyAtoms (number of heavy atoms, i.e. not hydrogen), nBonds (number of bonds, excluding bonds with hydrogen), FMF (complexity of a molecule), RotBFrac (fraction of rotatable bonds, excluding terminal bonds), XlogP (octanol-water partition coefficient by Wang et al. (1997)).

The model algorithm (PLS-DA) was coded in R (model stored as an R object and can be read by the R Model Reader Node), the applicability domain was calculated using Enalos Domain nodes.

Furthermore a web version was implemented, allowing to execute the model via a web browser: it is available through the COSMOS KNIME WebPortal (<http://knimewebportal.cosmostox.eu>).

Procedure Description

Biological Endpoint and Measurement:

Receptor binding: Liver X receptor (LXR)

Endpoint Value:

IC50: LXR β binding affinity values, converted into a binary variable: "active" (IC50 < 20 nM) and "inactives" (IC50 > 1000 nM)

Experimental System:

A mathematical QSAR model that predicts potential LXR binding by classifying chemicals as potential binders ("active") or not-binders ("inactive").

Quality/Acceptance Criteria

The QSAR model was developed according to the OECD principles for QSAR validation for regulatory purposes (OECD, 2006; OECD, 2007):

1. Defined endpoint: Liver X Receptor (LXR) binding potential
2. Defined algorithm: Partial Least Squares Discriminant Analysis (PLS-DA); six PaDEL descriptors (AMR, nHeavyAtom, nBonds, FMF, RotBTFrac, XLogP).
3. Defined applicability domain: leverage approach and similarity approach (based on Euclidean distances)
4. Internal and external validation: Cooper statistics provided for training and test sets.
5. Mechanistic interpretation: a posteriori mechanistic interpretation of the modelling descriptors.

3. Method Performance

Robustness of the Method

General performance measures:

Cooper statistics measuring goodness-of-fit:

Sensitivity = 0.85; Specificity = 0.84; Overall Accuracy = 0.85

FN = 6; TP = 34; TN = 32; FP = 6

FN: False Negative, i.e. active compound predicted as inactive; FP: False Positive, i.e. inactive compound predicted as active; TN: True Negative, i.e. inactive compound predicted as inactive; TP: True Positive, i.e. active compound predicted as active.

Test Compounds

The test set includes 19 compounds. Test set compounds were selected by *a priori* 30% random splitting of the whole dataset (N = 97) into a training set of 78 compounds and a test set of 19 compounds. The test set was not used for model development.

The test set is well represented by the training set. Training and test set are balanced in term of chemical structure and response.

Predictive capacity:

Cooper statistics measuring predictive performances:

Sensitivity = 0.70; Specificity = 0.89; Overall Accuracy = 0.79

FN = 3; TP = 7; TN = 8; FP = 1

FN: False Negative, i.e. active compound predicted as inactive; FP: False Positive, i.e. inactive compound predicted as active; TN: True Negative, i.e. inactive compound predicted as inactive; TP: True Positive, i.e. active compound predicted as active.

Applicability:

Applicability domain of the QSAR model is defined by two methodologies: the leverage approach and the similarity approach (based on Euclidean distances). The employed methodologies assess the "structural" distance of the target compound from the model training set.

Enalos Domain-Leverage (Melagraki et al., 2009; Afantitis et al., 2008) and Enalos Domain-Similarity (Melagraki et al., 2010; Afantitis et al., 2011) KNIME nodes are included in the workflow for the applicability domain analysis. Applicability domain thresholds (ADT) are defined for each approach (ADT for Leverage = 0.231; ADT for Similarity = 16.55). Compounds with leverage/similarity values greater than the specified thresholds are considered outside the applicability domain of the model and their predictions could be not reliable.

4. Discussion

Known experimental strengths and/or limitations of the method

- Freely available.
- Easy-to-use (user friendly interface - COSMOS KNIME WebPortal).
- No time-consuming (few minutes needed to run the workflow).
- Extendible (the QSAR model can be adapted, expanded and/or modified through the KNIME Analytics platform).
- Reproducible.
- The QSAR model can be applied in batch.
- The model only requires information on 2D chemical structure of the target compound (e.g., MOL, SMILES) to provide predictions.
- The modelling descriptors do not take into account 3D ligand-receptor interactions.
- Not suitable for inorganic compounds, polymers, mixtures or UVCB substances as well as for compounds falling outside the model applicability domain.
- Prediction of LXR binding potency, but not LXR activation.
-

Technical requirements

- The WebPortal version of the QSAR model can be run in a web browser without installation of the KNIME software
- The Desktop version requires installation of the KNIME Analytics Platform as well as additional nodes:
 - PaDEL Nodes
KNIME nodes for PaDEL can be downloaded from <http://padel.nus.edu.sg/software/padeldescriptor/>
 - R packages
To run the model a few additional packages have to be installed in R. If using a local R software: installation of caret, pls, e1071 packages with dependencies Without having R installed: go to you KNIME folder on local desktop, then follow into the folder `plugins\org.knime.ext.r.bin.win32.x86_2.15.2\R-Inst\bin`
Select the folder i386 or x64 - depending on the KNIME version - then double click on Rgui. When R starts, select Packages->Install packages from the top menu bar. Choose any of the mirror servers and accept clicking ok. From the long list of available packages choose: caret, pls and e1071.
 - Enalos nodes
Domain-Leverage and Domain-Similarity Enalos nodes available at: <http://www.novamechanics.com/knime.php>

Modifications of the method

A related QSAR model was developed using RDKit descriptors (instead of PaDEL descriptors). The RDKit-based QSAR model is implemented as a KNIME workflow and is available to be executed both in the COSMOS KNIME WebPortal and in the KNIME Analytics platform. Information on this model is available in the COSMOS Space (<http://cosmosspace.cosmostox.eu>).

Potential for future development

The workflow can be extended using additional or different modelling descriptors as well as different methodologies for applicability domain and reliability analysis.

Comparisons to other methods

Several *in silico* methodologies (e.g., structural alerts, 3D-QSAR, molecular modelling) are currently available to predict LXR binding, each characterised by strengths and limitations. Some models are published in the literature (e.g., Spreafico et al., 2010; Zhao et al., 2011; Salum et al., 2012, etc.) and/or have been implemented into software solutions (e.g., Virtual ToxLab™, Endocrine Disruptome (Kolšek et al., 2014), COSMOS-KNIME workflow for Nuclear Receptor Ligand Screening). In addition, within the COSMOS Project, a profiler based on a combination of different molecular modelling methodologies (namely, ensemble docking, e-Pharmacophore and fingerprints-based similarity) was developed and used in the context of SEURAT-1 Level 2 and Level 3 case studies (Fioravanzo et al., 2015). The complementary use of different *in silico* approaches would enhance the accuracy of predictions.

Additional Considerations

Additional details on the QSAR model and KNIME workflow as well as the user guides to run the workflow in KNIME Analytics platforms and in the KNIME Web Portal are available in COSMOS Space (<http://cosmospace.cosmostox.eu>).

5. Contact Details

Dr Simona Kovarich

Molecular modeling and QSAR
S-IN Soluzioni Informatiche S.r.l.
Corresponding author, Method developer / user
Via G. Ferrari n. 14
36100 Vicenza
Italy
email: simona.kovarich@s-in.it

Dr Anna Palczewska

Department of Computing
University of Bradford
Method developer / user
Bradford BD7 1DP
United Kingdom
email: a.palczewska@bradford.ac.uk

Ver n°	Date	Name	Type of change
1	03 November 2015	Dr Simona Kovarich	Data sheet created

6. Abbreviations & Definitions

ADT: Applicability Domain Threshold

AOP: Adverse Outcome Pathway

KNIME: Konstanz Information Miner, an open source data analytics, reporting and integration platform

IATA: Integrated Approaches to Testing and Assessment

ITS: Integrated Testing Strategy

LXR: Liver X Receptor

PLS-DA: Partial Least Squares - Discriminant Analysis

QMRF: QSAR Model Reporting Format

QSAR: Quantitative Structure-Activity Relationship

7. Bibliography

Method development:

- Afantitis A, Melagraki G, Koutentis PA, Sarimveis H, Kollias G (2011)
Ligand - based virtual screening procedure for the prediction and the identification of novel β -amyloid aggregation inhibitors using Kohonen Maps and Counterpropagation Artificial Neural Networks.
European Journal of Medicinal Chemistry 46, 497-508
- Afantitis A, Melagraki G, Sarimveis H, Koutentis PA, Markopoulos J, Igglessi-Markopoulou O (2008)
Development and Evaluation of a QSPR Model or the Prediction of Diamagnetic Susceptibility.
QSAR and Combinatorial Science 27, 432-436
- Melagraki G, Afantitis A, H. Sarimveis, P.A. Koutentis, Igglessi-Markopoulou O, Kollias G (2010)
In Silico Exploration for Identifying Structure–Activity Relationship of MEK Inhibition and Oral Bioavailability for Isothiazole Derivatives
Chemical Biology and Drug Design 76, 397–406
- Melagraki G., Afantitis A., Sarimveis H., Koutentis P.A., Kollias G., Igglessi-Markopoulou O. (2009)
Predictive QSAR workflow for the in silico identification and screening of novel HDAC inhibitors
Molecular Diversity 13 (3), 301-311

General references:

- Berthold MR., Cebron N., Dill F., Gabriel TR., Kötter T., Meinl T., Ohl P., Sieb C., Thiel K., Wiswedel B. (2007)
KNIME: The Konstanz Information Miner.
Studies in Classification, Data Analysis, and Knowledge Organization (GfKL 2007)
- Fioravanzo E, Bassan A, Kovarich S, Tsakovska I, Pajeva I, Palczewska A, Vitcheva V, Cronin MTD, Yang C, Worth A (2015)
Use of Molecular Modelling Approaches to Predict Potential Binding to Nuclear Receptors Involved in the Development of Liver Steatosis (i.e., LXR and PPAR γ)
In: Gocht T, Schwarz M (eds) Towards the replacement of in vivo repeated dose systemic toxicity testing. Volume 5, pp. 83-85
- Foundation Biographics Laboratory 3R, Basel/Switzerland ()
VirtualToxLab
Kolšek K, Mavri J, Dolenc MS, Gobec S, Turk S (2014)
Endocrine Disruptome - An Open Source Prediction Tool for Assessing Endocrine Disruption Potential through Nuclear Receptor Binding
Journal of Chemical Information and Modeling 54 (4), 1254–1267
- Landesmann B, Goumeou M, Munn S, Whelan M (2012)
Description of Prototype Modes-of-Action Related to Repeated Dose Toxicity
[JRC publication repository](#)
- OECD (2007)
Guidance document on the Validation of (Quantitative) Structure Activity Relationship [(Q)SAR] Models
OECD Environment Health and Safety Publications, Series on Testing and Assessment No. 69.
ENV/JM/MONO(2007)2
- OECD (2006)
Report on the Regulatory Uses and Applications in OECD Member Countries of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models in the Assessment of New and Existing Chemicals
OECD Environment Health and Safety Publications, Series on Testing and Assessment No. 58.
ENV/JM/MONO(2006)25
- Salum LB, Andricopulo AD, Honório KM (2012)
A fragment-based approach for ligand binding affinity and selectivity for the liver X receptor beta
Journal of Molecular Graphics and Modelling 32, 19
- Spreafico M, Smiesko M, Peristera O, Rossato G, Vedani A (2010)
Probing Small-Molecule Binding to the Liver-X Receptor: A Mixed-Model QSAR Study
Molecular Informatics 29, 27

- Yap CW (2011)
PaDEL-Descriptor: An open source software to calculate molecular descriptors and fingerprints
Journal of Computational Chemistry 32 (7), 1466-1474
- Zhao W, Gu Q, Wang L, Ge H., Li J, Xu J (2011)
Three dimensional pharmacophore modeling of liver-X receptor agonists
Journal of Chemical Information and Modeling 51, 2147

Last update: 15 November 2015

DB-ALM Method summary n° 175 : PPAR full agonists' potency predictor

Systemic Toxicity

PPAR full agonists' potency predictor is a three dimensional QSAR (Quantitative Structure-Activity Relationships) model for prediction of chemical compounds' potency (transactivation activity, pEC₅₀) to induce PPAR γ -mediated expression of target reporter gene. The model predicts transactivation activity using calculated electrostatic, acceptor and hydrophobic similarity indices as 3D descriptors, generated by CoMSIA (Comparative Molecular Similarity Indices Analysis) 3D QSAR approach.

1. General Information

Corresponding author(s):

Dr Ivanka Tsakovska, Institute of Biophysics and Biomedical Engineering, Bulgarian Academy of Sciences (BAS) Bulgaria

Date of data sheet creation: 16 November 2015

Last version number: 1

Last update: 16 November 2015

Section **No.5 Contact Details** provides a complete list of the authors and method developers.

The method summary presented here describes a mathematical model.

In the context of this document, the term "method" in the paragraph headings refers to the model.

Proprietary and/or Confidentiality Issues:

Public

Status

Development

Development completed. PPAR γ full agonists' potency predictor was developed as a part of COSMOS project (Integrated *In Silico* Models for the Prediction of Human Repeated Dose Toxicity of Cosmetics to Optimise Safety), within the SEURAT-1 Research Initiative (Safety Evaluation Ultimately Replacing Animal Testing 1). More information on the project is available on its dedicated website: <http://www.cosmostox.eu>.



2. Method Definition

Purpose of the Method

TYPE OF TESTING	: Screening, Part of an integrated testing strategy
LEVEL OF TOXICITY ASSESSMENT	: Toxic potential

Context of use

Toxicity-induced liver injury involves a spectrum of pathological conditions, including NAFLD (hepatic steatosis / steatohepatitis), which may progress toward cirrhosis and hepatocellular carcinoma. Chemical hazard identification and characterisation at early stages, such as expression of NAFLD biomarkers, are central in a risk assessment strategy focused on repeated dose hepatotoxicity. The developed 3D QSAR model is based on an Adverse Outcome Pathway (AOP) (Al Sharif et al., 2014), which integrated experimental evidence for an array of PPAR γ target proteins. Since their overexpression synergistically drives disease development and progression, the *in silico* prediction of PPAR γ ligands' transactivation activity is suggested as mechanistically justified rational for the screening and prioritisation for further testing of potentially prosteatotic chemicals.

Scientific Principle of the Method

PPAR γ activation by full agonists has been outlined as one of the possible molecular initiating events for mode of action leading to non-alcoholic fatty liver disease (NAFLD) (Al Sharif et al., 2014). Agonist binding involves non-covalent protein-ligand interactions leading to PPAR γ activation and triggering target protein expression. The development of a 3D QSAR model based on this biological response involved:

- i. selection of full PPAR γ agonists that are expected to share common mode of action and receptor binding mode
- ii. generation and alignment of potentially bioactive conformers
- iii. calculation of 3D properties, relevant to receptor binding
- iv. establishment of predictive quantitative structure-activity relationship using CoMSIA 3D QSAR approach

The experimental data reported in the literature used for the selected endpoint has been obtained from Luciferase reporter gene assay. The last is relevant to the OECD conceptual framework, presenting the most promising assays to detect and characterise chemical effects on the PPAR signaling pathway and is among the *in vitro* assays that are going to be considered for incorporation into new or existing Test Guidelines for the detection of endocrine disrupting chemicals after refinement and validation (OECD, ENV/JM/MONO(2012)23).

Model Description

An initial dataset (<http://biomed.bas.bg/qsarmm/>) with structural and biological (binding affinity, IC₅₀; transactivation activity, EC₅₀; relative efficacy, %max) data for 439 PPAR γ agonists (full and partial) was subjected to refinement by: (i) data gaps removal; (ii) selection of full agonists (relative efficacy \geq 70%) avoiding duplicates and data uncertainties; (iii) stereochemistry consideration – with a preference for S stereoisomers when the potency of racemic mixtures was reported. For the selected 170 full agonists the negative logarithm of the EC₅₀ data was used as dependent variable in the model development.

The 170 structures of the selected compounds were generated based on structurally similar scaffolds of Protein Data Bank (PDB) ligands and were aligned by pharmacophore-based docking procedure (Tsakovska et al., 2014) in the ligand-binding domain of PPAR γ using the molecular modelling (MM) platform MOE v. 2014.0901 (CCG Inc.). The conformers were selected based on docking scores and visual inspection versus template structure (ligand UNT from PDB ID 3IA6; pEC₅₀ = 7.886, 103% relative efficacy). For these compounds CoMSIA steric, electrostatic, donor, acceptor and hydrophobic fields were calculated in MM software suite SYBYL-X v. 2.1 (Certara USA, Inc.). Applicability domain outliers and response outliers were excluded. The final modelling set consisted of 122 compounds (83 – training, 39 – test set). Multiple 3D QSAR models were obtained. Their internal and external predictivity were assessed through leave-one-out (LOO) cross-validation and potency prediction of a test set, respectively.

CoMSIA electrostatic, acceptor and hydrophobic fields were used to correlate 3D structure variability with differences in transactivation activity through Partial Least Squares (PLS) statistical analysis. The developed 3D QSAR model is able to predict the potency of full PPAR γ agonists.

Procedure Description

Biological Endpoint and Measurement:

This method is an *in silico* model, which provides information on the following, biologically relevant, endpoint:

Ligand-induced transactivation of peroxisome proliferator-activated receptor gamma (PPAR γ)

Endpoint Value:

EC₅₀

Experimental System:

In silico mathematical model to predict the potential of chemicals to bind and activate PPAR γ 1 and PPAR γ 2, jointly referred to as PPAR γ

3. Method Performance

Robustness of the Method

General performance measures:

The robustness of the model was evaluated by LOO cross-validation basing on the cross-validated coefficient q_{CV}^2 , the optimal number of principle components N_{opt} , and the cross-validated standard error of prediction (SEP_{CV}).

$$LOO\ q_{CV}^2 = 0.610$$

$$N_{opt} = 7$$

$$SEP_{CV} = 0.505$$

Test Compounds

The modelling subset of 122 full PPAR γ agonists utilised in this study comprised different homologous series retrieved from 15 literature sources: eight series of chemicals contain a PPAR γ ligand with a crystal structure deposited in the PDB, one contains a PPAR α ligand, and six do not contain resolved PDB structures (the literature sources and training/test set assignment for the compounds used in 3D QSAR modelling are available in the dataset description at <http://biomed.bas.bg/qsarmm/>). The training set includes 83 compounds with slight prevalence of human over animal *in vitro* data, wide range of pEC₅₀ data and consisting of structures from all collected series of compounds. The test set consists of 39 full PPAR γ agonists, within the applicability domain of the model.

Predictive capacity:

The predictive power of the obtained model was evaluated by external validation. The predictive correlation coefficient r_{pred}^2 was calculated with training ($n_{training\ set}$) to test set ($n_{test\ set}$) ratio approx. 2:1.

Statistical parameters:

$$n_{training\ set} = 83$$

$$n_{test\ set} = 39$$

$$r_{pred}^2 = 0.552$$

Applicability:

The model is applicable to full PPAR γ agonists. The applicability domain outliers were identified on the basis of the calculated CoMSIA fields using the "extent of extrapolation" approach (Tropsha et al., 2003; Netzeva et al., 2005) as implemented in Enalos domain leverage node (Melagraki et al., 2009) in the KNIME Analytics Platform (Berthold et al., 2007).

4. Discussion

Ethical issues and considerations for 3R's impact

The combination of *in vitro* experimental data and *in silico* modelling follows the main principles of the 3Rs. Compared to previously published 3D QSAR pEC₅₀-based models, this one is significantly enriched with data from chemicals' testing on human cell lines (human to animal cell line ratio is 0.649/1.147 for the training/test sets) thus underlining the effort towards humane and reliable safety evaluation.

Known experimental strengths and/or limitations of the method

The model has straightforward mechanistic interpretation. It predicts ligand-induced *in vitro* transactivation (EC₅₀) since it is capable to reflect the agonistic activity of the compounds. The EC₅₀ covers the complex cascade of receptor binding and activation followed by the downstream molecular events triggering gene expression, therefore it is an appropriate *in vitro* experimental model of the

quantitative relationship between the MIE (PPAR γ activation) and the earliest downstream key event – increased synthesis of target proteins. Although complex in its nature and thus challenging to be modelled, EC₅₀ may reflect, in a more complete manner, the mechanism behind the particular pathology.

The quality of the model may be limited by intra- and inter-laboratory variations in the experimental settings for measuring the test compounds' transactivation activity and in the selection of a reference compound for calculating their relative efficacy.

The application of the model requires availability of molecular modelling specialised software.

Technical requirements

Computer system capable of running the MM platforms MOE v. 2014.0901 (CCG Inc.) and SYBYL-X v. 2.1 (Certara USA, Inc.).

Potential for future development

Additional external validation could be performed based on a test set consisting of experimentally observed prostatic chemicals.

Comparisons to other methods

The 3D QSAR model of PPAR γ agonism relies on the bioactive conformers, providing more comprehensive mechanistical understanding and interpretation of the molecular events.

Using potency instead of binding affinity data reflects, in a more complete manner, the mechanism behind the considered pathology.

The developed 3D QSAR model, involving electrostatic, acceptor and hydrophobic CoMSIA fields is as a good improvement over the previous ones published in the scientific literature, based on:

- i. the largest and structurally diverse training set (83 full agonists)
- ii. reasonable training to test set ratio (83/39, 2:1)
- iii. reasonable pEC₅₀ range (3.7/2.6 log units for the training/test sets)
- iv. reporting the full range of statistical parameters necessary to apply the model for predictive purposes

5. Contact Details

Dr Merilin Al Sharif

Department of QSAR & Molecular Modelling
Institute of Biophysics and Biomedical Engineering, Bulgarian Academy of Sciences (BAS)
Method developer / user
105 G. Bontchev Str
Sofia 1113
Bulgaria
email: merilin.al@biomed.bas.bg
telephone: +359 2 9793647

Petko Alov

Department of QSAR & Molecular Modelling
Institute of Biophysics and Biomedical Engineering, Bulgarian Academy of Sciences (BAS)
Method developer / user
105 G. Bontchev Str
Sofia 1113
Bulgaria
email: petko.alov@biophys.bas.bg
telephone: +359 2 9793647

Dr Ivanka Tsakovska

Department of QSAR & Molecular Modelling
Institute of Biophysics and Biomedical Engineering, Bulgarian Academy of Sciences (BAS)
Corresponding author, Method developer / user

105 G. Bontchev Str
Sofia 1113
Bulgaria
email: ITsakovska@biomed.bas.bg
telephone: +359 2 9793605

Dr Ilza Pajeva

Department of QSAR & Molecular Modelling
Institute of Biophysics and Biomedical Engineering, Bulgarian Academy of Sciences (BAS)
Method developer / user
105 G. Bontchev Str
Sofia 1113
Bulgaria
email: pajeva@biomed.bas.bg
telephone: +359 2 9793605

Ver n°	Date	Name	Type of change
1	16 November 2015	Dr Ivanka Tsakovska	Data sheet created

6. Abbreviations & Definitions

AOP: Adverse outcome pathway

CoMSIA: Comparative molecular similarity indices analysis

COSMOS: EU Integrated Project. COSMOS developed an integrated suite of computational tools to predict the effects of long-term exposure to chemicals in humans based on in silico methods. Website: <http://www.cosmostox.eu/>

LOO: Leave-one-out

MM: Molecular modelling

NAFLD: Non-alcoholic fatty liver disease

N_{opt}: Optimal number of PLS components

OECD: Organization for Economic Cooperation and Development

pEC₅₀: Negative logarithm of the half maximal effective concentration

PLS: Partial least squares

PPAR_γ: Peroxisome proliferator-activated receptor γ

q_{cv}²: Squared cross-validated correlation coefficient

r_{pred}²: Squared predictive correlation coefficient

SEP_{cv}: Cross-validated standard error of estimation

SEURAT 1: Safety Evaluation Ultimately Replacing Animal Testing 1

7. Bibliography

Method development:

- Al Sharif M., Tsakovska I., Pajeva I., Alov P., Fioravanzo E., Bassan A., Kovarich S., Yang C., Mostrag-Szlichtyng A., Vitcheva V., F, Worth A.P., Richarz A.-N., Cronin MTD. (2016) The application of molecular modelling in the safety assessment of chemicals: A case study on ligand-dependent PPAR_γ dysregulation. *Toxicology In press*, <http://dx.doi.org/10.1016/j.tox.2016.01.009>
- Al Sharif, M., Alov, P., Vitcheva, V., Pajeva, I., Tsakovska, I. (2014)

Modes-of-Action Related to Repeated Dose Toxicity: Tissue-Specific Biological Roles of PPAR γ Ligand-Dependent Dysregulation in Nonalcoholic Fatty Liver Disease
PPAR Research 2014, 432-647

- Berthold MR., Cebron N., Dill F., Gabriel TR., Kötter T., Meini T., Ohi P., Sieb C., Thiel K., Wiswedel B. (2007)
 KNIME: The Konstanz Information Miner.
Studies in Classification, Data Analysis, and Knowledge Organization (GfKL 2007)
- COSMOS (2016)
<http://www.cosmostox.eu/>
- Fratev F., Tsakovska I., Al Sharif M., Mihaylova E., Pajeva I. (2015)
 Structural and dynamical insight into PPAR γ antagonism: in silico study of the ligand-receptor interactions of non-covalent antagonists.
International Journal of Molecular Sciences 16(4), 15405-15424
- Melagraki G., Afantitis A., Sarimveis H., Koutentis P.A., Kollias G., Igglessi-Markopoulou O. (2009)
 Predictive QSAR workflow for the in silico identification and screening of novel HDAC inhibitors
Molecular Diversity 13 (3), 301-311
- Molecular Operating Environment (MOE) (2015)
 version 2014.091; Chemical Computing Group Inc., 2015, <http://www.chemcomp.com>.
- Netzeva TI., Worth A., Aldenberg T., Benigni R., Cronin M.T., Gramatica P., Jaworska J.S., Kahn S., Klopman G., Marchant C.A., Myatt G., Nikolova-Jeliazkova N., Patlewicz G.Y., Perkins R., Roberts D., Schultz T., Stanton D.W., van de Sandt J.J., Tong W., Veith G., Yang C. (2005)
 Current Status of Methods for defining the Applicability of (Quantitative) Structure-Activity Relationships.
Alternatives to Laboratory Animals (ATLA) 33, 155-173
- OECD (2012)
 OECD Guidance Document No. 178: Detailed review paper on the state of the science on Novel in vitro and in vivo screening and testing methods and endpoints for evaluating endocrine disruptors. (Authors: LeBlanc, G., Norris, D., and Kloas, W.)
OECD Environmental Health and Safety Publications 178
- PPAR γ ligands' dataset (2016)
<http://biomed.bas.bg/qsarmm/>
- SYBYL-X (2013)
 version 2.1, Tripos International, 2013, <https://www.certara.com/>
- Tropsha, A., Gramatica, P., Gombar, V. (2003)
 The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models
QSAR and Combinatorial Science 2, 69–77
- Tsakovska, I., Al Sharif, M., Alov, P., Diukendjieva, A., Fioravanzo, E., Cronin, M., Pajeva, I. (2014)
 Molecular Modelling Study of the PPAR γ Receptor in Relation to the Mode of Action/Adverse Outcome Pathway Framework for Liver Steatosis.
International Journal of Molecular Sciences 15, 7651–7666

Last update: 15 November 2015

DB-ALM Method summary n° 176 : Parallel Artificial Membrane Permeability Assay (PAMPA) logPm predictor

Systemic Toxicity

PAMPA logPm predictor is a mathematical model for prediction of chemical compounds' permeability in Parallel Artificial Membrane Permeability Assay (PAMPA), which is assumed to be a practical estimator of passive gastro-intestinal absorption (GIA) of low molecular weight compounds. The model predicts permeability coefficient Pm using *in silico* calculated distribution coefficient (logD) and the ratio of the Topological Polar Surface Area (TPSA) to Molecular Weight (MW) as descriptors.

1. General Information

Corresponding author(s):

Petko Alov, Institute of Biophysics and Biomedical Engineering, Bulgarian Academy of Sciences (BAS) Bulgaria

Date of data sheet creation: 11 November 2015

Last version number: 1

Last update: 11 November 2015

Section **No.5 Contact Details** provides a complete list of the authors and method developers.

The method summary presented here describes a mathematical model.

In the context of this document, the term "method" in the paragraph headings refers to the model.

Proprietary and/or Confidentiality Issues:

Public

Status

Development

Development completed. The model was developed in the FP7 COSMOS Project, which was part of the Research Initiative - SEURAT-1 - Towards the Replacement of *In Vivo* Repeated Dose Systemic Toxicity Testing (2011-2015). More information on the COSMOS project is available on its dedicated website:

<http://www.cosmostox.eu>



The model is implemented as a KNIME workflow, available on the COSMOS project COSMOS Space website:

<http://cosmosspace.cosmostox.eu>

A web version of the model was also implemented, allowing to execute the model via a web browser. It is available through the COSMOS KNIME WebPortal:

<http://knimewebportal.cosmostox.eu>

A tutorial is available at:

<https://www.youtube.com/watch?v=WJLAz9GnvcM/>

2. Method Definition

Purpose of the Method

TYPE OF TESTING : Screening

Context of use

The estimation of bioavailability after oral and dermal administration is of key importance for example in prediction of the chronic toxicity of cosmetic-related ingredients: whilst cosmetics are usually applied dermally, the majority of available repeated dose toxicity data are obtained from oral administration.

The method predicts a parameter useful in further predictions or exposure assessment. **PAMPA logP_m predictor** is a mathematical QSAR model for prediction of chemical compounds' permeability in Parallel Artificial Membrane Permeability Assay (PAMPA), expressed as logP_m, which is assumed to be a practical estimator of passive gastro-intestinal absorption (GIA) of low molecular weight compounds.

Scientific Principle of the Method

The implemented QSAR mode was developed using a data set of 276 compounds from the Database of Double-Sink PAMPA logP₀, logP_m at pH6.5 and logP_m at pH7.4 (Avdeef A., 2012). It predicts permeability coefficient P_m using *in silico* calculated distribution coefficient (logD) and the ratio of the Topological Polar Surface Area (TPSA) to Molecular Weight (MW) as descriptors.

Data from Avdeef (2012) consisted of 290 entries for PAMPA logP_m, obtained by PAMPA-DS (double-sink) method. The data collection was subjected to removal of entries studied with variations of the method, chemical structure retrieval (by means of NCI/CADD Chemical Identifier Resolver and NCBI PubChem project), mixtures removal and desalting (CDK toolbox as implemented in KNIME Analytics Platform). CAS numbers were retrieved for the majority of the compounds. The final number of compounds used for modelling was 276.

The model is devised as an improvement of the model reported by Nakao and co-workers (2009). The Nakao et al. (2009) model used logP, |pK_a - pH|, and Topological Polar Surface Area (TPSA) as descriptors, the present one uses the logD and TPSA/MW ratio. The rationale for changing the descriptors was that (a) log D ≈ logP - |pK_a - pH|, and (b) molecular weight (MW) is used as an estimator of the total surface to account for the ratio of polar to total surface area.

With a view to the free and open implementation of the models and lack of such tools for logD estimation, two implementations of the model were produced based on logD estimations readily obtainable through free online services ChemSpider.com (calculated by ACD/Labs tools) and Chemicalize.org (calculated by ChemAxon tools).

Model Description

The KNIME analytics Platform was used for the model implementation.

Two implementations of the model were produced based on logD estimations readily obtainable through free online services:

- ChemSpider.com (calculated by ACD/Labs tools)
- Chemicalize.org (calculated by ChemAxon tools)

The model development included removal of the applicability domain and response outliers (Gramatica, 2007), and final expressions are:

ACD/Labs-calculated-logD implementation:

$$\log P_m = -2.945(\pm 0.228) + 0.600(\pm 0.046)\log D_{pH7.4} - 7.655(\pm 0.811)TPSA/MW$$

$$n=246, r^2_{adj}=0.734, SEE=1.108, F=338.861$$

ChemAxon-calculated-logD implementation:

$$\log P_m = -2.280(\pm 0.240) + 0.494(\pm 0.050)\log D_{pH7.4} - 10.063(\pm 0.844)TPSA/MW$$

$$n=247, r^2_{adj}=0.711, SEE=1.164, F=303.713$$

Procedure Description

Biological Endpoint and Measurement:

Artificial Membrane Permeability Assay (PAMPA) is assumed to be a practical estimator of passive

gastro-intestinal absorption (GIA) of low molecular weight compounds

Endpoint Value:

LogP_m - membrane permeability coefficient

Experimental System:

A mathematical QSAR model for prediction of chemical compounds' permeability in Parallel Artificial Membrane Permeability Assay (PAMPA)

3. Method Performance

Test Compounds

Curated data from Avdeef (2012) consisted of 276 compounds: commercial drugs, pesticides, several derivatives of commercial drugs, and several other chemical compounds.

Predictive capacity:

Predictive capacity of the model was assessed by leave-one-out (LOO) cross-validation and by external validation by means of splitting of the dataset into training (~80%) and test (~20%) groups.

ACD/Labs-calculated-logD implementation:

LOO q²=0.729, external validation q²=0.696 (training set 196 / test set 50)

ChemAxon-calculated-logD implementation:

LOO q²=0.705, external validation q²=0.722 (training set 197 / test set 50)

Applicability:

Applicability domain of the model was assessed by the extent of extrapolation approach (based on the calculation of the leverage h_i for each chemical compound) as implemented in the KNIME community node developed by NovaMechanics.

4. Discussion

Known experimental strengths and/or limitations of the method

Simplicity, straightforward mechanistic interpretation, implementation in versatile open-source KNIME Analytics Platform.

Technical requirements

Computer system capable of running KNIME Analytics Platform.
The COSMOS KNIME WebPortal version can be executed through a web browser (<http://knimewebportal.cosmostox.eu>).

Modifications of the method

Two modifications are described in the previous sections.

Potential for future development

Other *in silico* estimations of logD could be studied as potential descriptor in the model.

Comparisons to other methods

The present model is built on an extended data set of 276 compounds (compared to max. 97 compounds

in available literature, Fujikawa et al., 2007). The endpoint Pm was obtained by the most versatile variation of the PAMPA method – double-sink PAMPA. The model is applicable to compounds with much broader range of hydrophobicity as compared to similar models (Fujikawa et al., 2007; Nakano et al., 2012).

5. Contact Details

Petko Alov

Department of QSAR & Molecular Modelling
Institute of Biophysics and Biomedical Engineering, Bulgarian Academy of Sciences (BAS)
Corresponding author, Method developer / user
105 G. Bontchev Str
Sofia 1113
Bulgaria
email: petko.alov@biophys.bas.bg
telephone: +359 2 9793647

Dr Ivanka Tsakovska

Department of QSAR & Molecular Modelling
Institute of Biophysics and Biomedical Engineering, Bulgarian Academy of Sciences (BAS)
Method developer / user
105 G. Bontchev Str
Sofia 1113
Bulgaria
email: ITsakovska@biomed.bas.bg
telephone: +359 2 9793605

Dr Ilza Pajeva

Department of QSAR & Molecular Modelling
Institute of Biophysics and Biomedical Engineering, Bulgarian Academy of Sciences (BAS)
Method developer / user
105 G. Bontchev Str
Sofia 1113
Bulgaria
email: pajeva@biomed.bas.bg
telephone: +359 2 9793605

Ver n°	Date	Name	Type of change
1	11 November 2015	Petko Alov	Data sheet created

6. Abbreviations & Definitions

logD: 1-octanol / water distribution coefficient at a particular pH

logP: 1-octanol / water partition coefficient

logPm: membrane permeability coefficient

LOO: leave-one-out cross-validation

MW: molecular weight

PAMPA: parallel artificial membrane permeability assay

pKa: acid dissociation constant

SEE: standard error of estimation

TPSA: topological polar surface area

7. Bibliography

Method development:

- Avdeef A. (2012)
Absorption and Drug Development. Solubility, Permeability, and Charge State, 2nd Edition
John Wiley; New York, Chichester., pp. 698
- Fujikawa M., Nakao K., Shimizu R. and Akamatsu M. (2007)
QSAR study on permeability of hydrophobic compounds with artificial membranes
Bioorganic & medicinal chemistry 15, 3756-3767
- Gramatica P. (2007)
Principles of QSAR models validation: internal and external.
QSAR and Combinatorial Science 26, 694-701
- Nakao K., Fujikawa M., Shimizu R. and Akamatsu M (2009)
QSAR application for the prediction of compound permeability with in silico descriptors in practical use
Journal of computer-aided molecular design 23, 309-319

Last update: 15 November 2015

DB-ALM Method summary n° 177 : *In silico* screening for potential nuclear receptor ligands and alerts towards hepatosteatosis

Systemic Toxicity

A workflow based on the computational structural alerts to screen for nuclear receptor agonists, containing 214 structural features, recorded computationally as SMARTS strings. They can be used for grouping and screening during drug development and risk assessment and provide knowledge to anchor adverse outcome pathways (AOPs).

1. General Information

Corresponding author(s):

Prof. Mark Cronin, Liverpool John Moores University

Date of data sheet creation: 12 November 2015

Last version number: 1

Last update: 12 November 2015

Proprietary and/or Confidentiality Issues:

None. The model is publicly available.

Status

Development

Development completed. Work carried out at Liverpool John Moores University with support from the FP7 COSMOS Project (Integrated *In Silico* Models for the Prediction of Human Repeated Dose Toxicity of Cosmetics to Optimise Safety), within the SEURAT-1 (Safety Evaluation Ultimately Replacing Animal Testing 1) cluster (2011-2015). More information on the COSMOS project is available on its dedicated website:

<http://www.cosmostox.eu>



The *in silico* workflow contains a number of structural alerts coded into SMARTS patterns and translated into KNIME (<http://www.knime.org/>) automated workflows as part of the COSMOS project. The KNIME workflow is available via COSMOS Space:

<http://cosmospace.cosmostox.eu>

A web version of the model was also implemented, allowing to execute the model via a web browser. It is available through the COSMOS KNIME WebPortal:

<http://knimewebportal.cosmostox.eu>

A tutorial is available at:

<https://www.youtube.com/watch?v=ggkU6IZfDfY>

2. Method Definition

Purpose of the Method

TYPE OF TESTING	: Screening, Part of an integrated testing strategy
LEVEL OF TOXICITY ASSESSMENT	: Toxic potential

Context of use

Many pathways in toxicology can be traced back to nuclear receptors (NR) as initial target. One notable example is endocrine disruption, another the development of fatty liver (hepatosteatosis) associated with chronic exposure to the agonists of some nuclear receptors, e.g. Retinoic Acid Receptor, Retinoid X receptor, Liver X Receptor and Peroxisome Proliferator-Activated Receptor (Moya et al 2010).

For the purpose of this model the structural and physico-chemical features of NR ligands have been investigated using data from ChEMBL and the Protein Data Bank (PDB). The information obtained was implemented into a KNIME workflow (see Steinmetz et al 2015) and can be used for screening for potential NR ligands during drug development and risk assessment, which can support hazard identification, category formation for read-across and provide knowledge to anchor adverse outcome pathways (AOPs).

Scientific Principle of the Method

Nuclear receptor activation by xenobiotics has been linked to the onset of many different toxicities such as endocrine disruption and hepatosteatosis. Therefore it is an important parameter in considering the effects of chronic chemical exposure.

A workflow to screen for potential NR ligands and alerts towards hepatosteatosis has been developed by the authors. Chemical descriptors required for receptor activation were implemented along with key structural features which have been coded into SMARTS (SMiles ARbitrary Target Specification) strings and translated into a KNIME workflow. SMARTS pattern allows for a definition of a sub-structure within a molecule, which is then used to check whether a chemical structure contains any of the pre-defined fragments.

The NRs associated with the onset of hepatosteatosis were identified, according to Moya et al (2010). They are listed in **Table 1**. In the developed screening workflow that identifies potential agonists of all NR of interest, those linked to hepatosteatosis are flagged with a special alert in the workflow output.

In total the following NR are included in the workflow: Aryl Hydrocarbon Receptor (AHR), Androgen Receptor (AR), Estrogen Receptor (ER), Farnesoid X Receptor (FXR), Glucocorticoid Receptor (GR), Liver X Receptor (LXR), Peroxisome Proliferator-Activated Receptor (PPAR), Progesterone Receptor (PR), Pregnane X Receptor (PXR), Retinoic Acid Receptor (RAR)/ Retinoid X receptor (RXR), Thyroid Receptor (THR), Vitamin D Receptor (VDR).

Table 1. Nuclear receptors associated with hepaosteatois and abbreviations as defined by Mellor et al. (2015).

Nuclear receptor name	Abbreviation
Aryl Hydrocarbon Receptor	AHR
Constitutive Androstane Receptor	CAR
Estrogen Receptor	ER
Farnesoid X Receptor	FXR
Glucocorticoid Receptor	GR
Liver X Receptor	LXR
Peroxisome Proliferator-Activated Receptor	PPAR
Pregnane X Receptor	PXR
Retinoic Acid Receptor	RAR
Retinoid X receptor	RXR

Model Description

A computational structural alerts based workflow to screen for potential nuclear receptor agonists and alerts towards hepatosteatosis was developed using data obtained from the ChEMBL database of active molecules and the protein data bank (PDB) (Mellor et al., 2016). The information was used to identify common structural and chemical features and the chemical space of known NR agonists and the essential ligand-binding interactions that they form in order to bind to the nuclear receptor and cause

activation.

The identified rules were implemented in a KNIME workflow. The workflow uses a combination of molecular fragments and other relevant chemical features. In total 214 structural features were recorded computationally as SMARTS strings.

Source Data

Data was obtained from the ChEMBL database and the protein data bank (PDB).

Extracting data from ChEMBL

The ChEMBL (ChEMBL, 2015) database is a database of bioactive molecules comprising over 1.5 million compounds and over 9,000 biological targets. Data from compounds with K_i , K_d , AC_{50} and EC_{50} values towards the NRs of interest were downloaded, combined and sorted by their pChEMBL values. The pChEMBL value is an approach to standardise different types of activity values (Bento et al., 2013). Every compound with a value not less than **five** was regarded as active compound. Common substructures were compared to ligand-protein-interaction information derived from PDB (PDB, 2015).

Analysis of agonistic ligands using the PDB

The PDB was searched for human NR structures. The human structures were then investigated regarding their ligand-protein-interaction within PyMOL 1.3 (PyMOL, 2014) and studying the linked publications. Important structural features for ligand-protein-interactions, e.g. functional groups forming hydrogen bonds, were coded into SMARTS strings, listed in detail in the workflow description in COSMOS Space (<http://cosmospace.cosmostox.eu>) and in Mellor et al (2016).

Physico-chemical space

The physico-chemical space was investigated using the CDK node for molecular properties within KNIME 2.9.4 (incl. community contributions). Different types of descriptors (vertex adjacency information magnitude (VAIM) for structural complexity, number of rotational bonds (RB) for flexibility, molecular weight (MW) for molecular size and Log K_{OW} (XLogP) for lipophilicity), were calculated.

Chemical features identified

After studying the data obtained from ChEMBL and PDB, the common structural features of the NR agonists were identified along with the essential functional groups needed for ligand binding to each NR, e.g. hydroxyl groups needed to form hydrogen bonds with key amino acid residues on the NR.

Building rules for the screening workflow

Rules based on physico-chemical properties and structural features were established. The structural features and essential functional groups were coded into SMARTS strings. The rules and SMARTS strings applied for each NR are listed in detail in the workflow description in COSMOS Space (<http://cosmospace.cosmostox.eu>).

KNIME implementation

A KNIME workflow executing the rules was built based on KNIME community contribution nodes. It is checking the physico-chemical space (CDK node for calculating molecular properties) and for important structural features (Indigo substructure matcher). Data processing is executed with standard KNIME nodes.

The NR ligands which are associated with hepatosteatosis get an extra alert in addition to the NR prediction. It shall be pointed out that many NR ligands are regarded as promiscuous, i.e. bind more than one receptor. Thus multiple predictions are likely, and hence shown by employing the grouping node. As the workflow is thought to be an initial screening tool for large datasets, only ligands with at least one NR prediction are shown in the results. In case no prediction/alert is triggered, a message saying “no potential NR ligands found” will pop-up, so that the user knows that the workflow was executed properly.

The methods used to create the workflow are described in Steinmetz et al (2015) and Mellor et al (2016).

An overview of the KNIME workflow is shown in **Figure 1**, details on substructure checking within the workflow in **Figure 2**.

The model was implemented as KNIME workflow web version, an easy-to-use user-interface allowing to execute the model via a web browser. It is available through the COSMOS KNIME WebPortal:

<http://knimewebportal.cosmostox.eu>

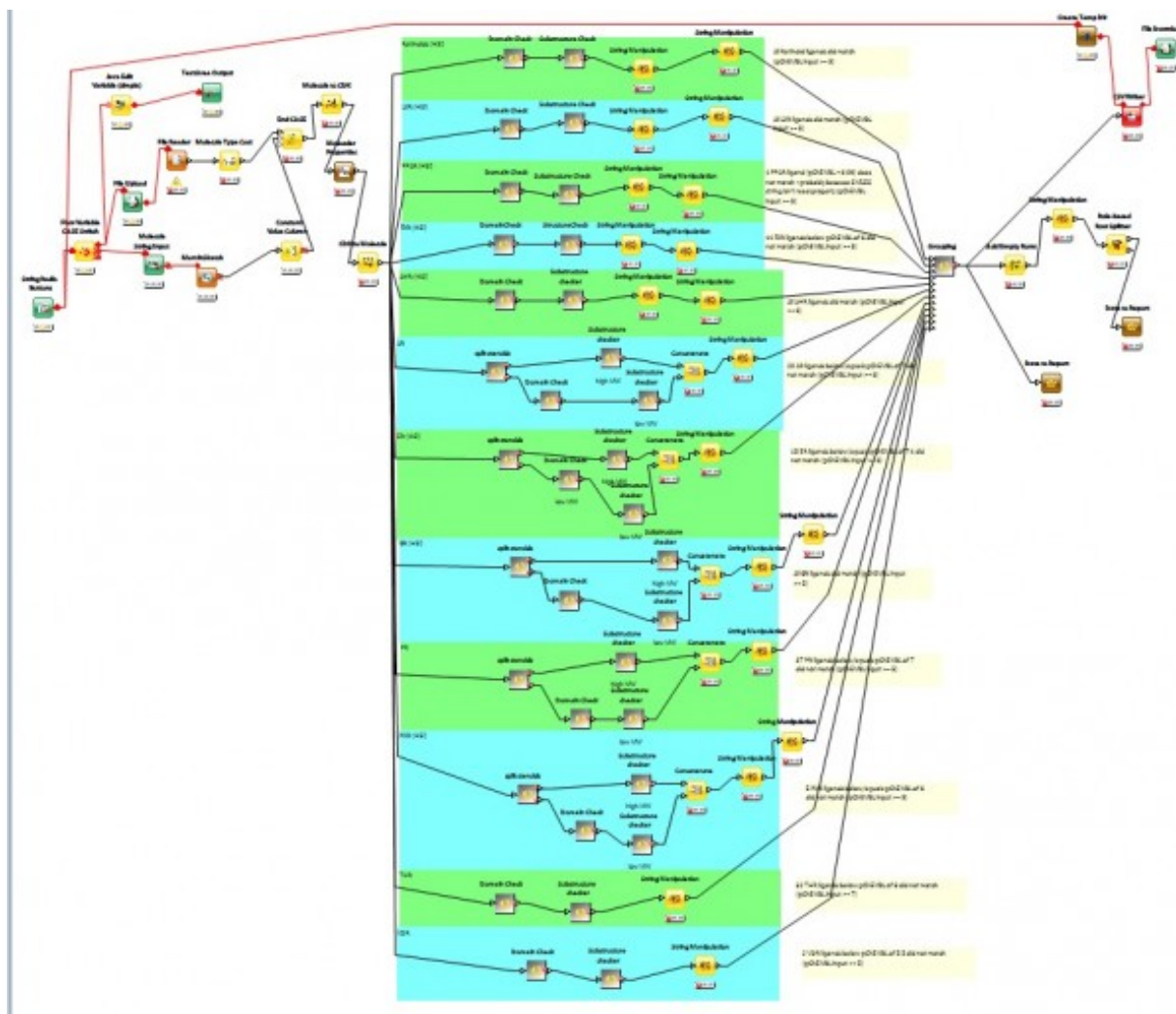


Figure 1: Workflow overview

The KNIME workflow is composed of twelve sub-workflows, one for each NR. In each workflow the information on relevant substructures (SMARTS strings of scaffolds and functional groups) and physico-chemical features, e.g. ranges for molecular weight, log P, vertex adjacency information magnitude, topical polar surface area, number hydrogen bond donors and rotational bonds, is checked and if applicable the compound is flagged as a potential ligand.

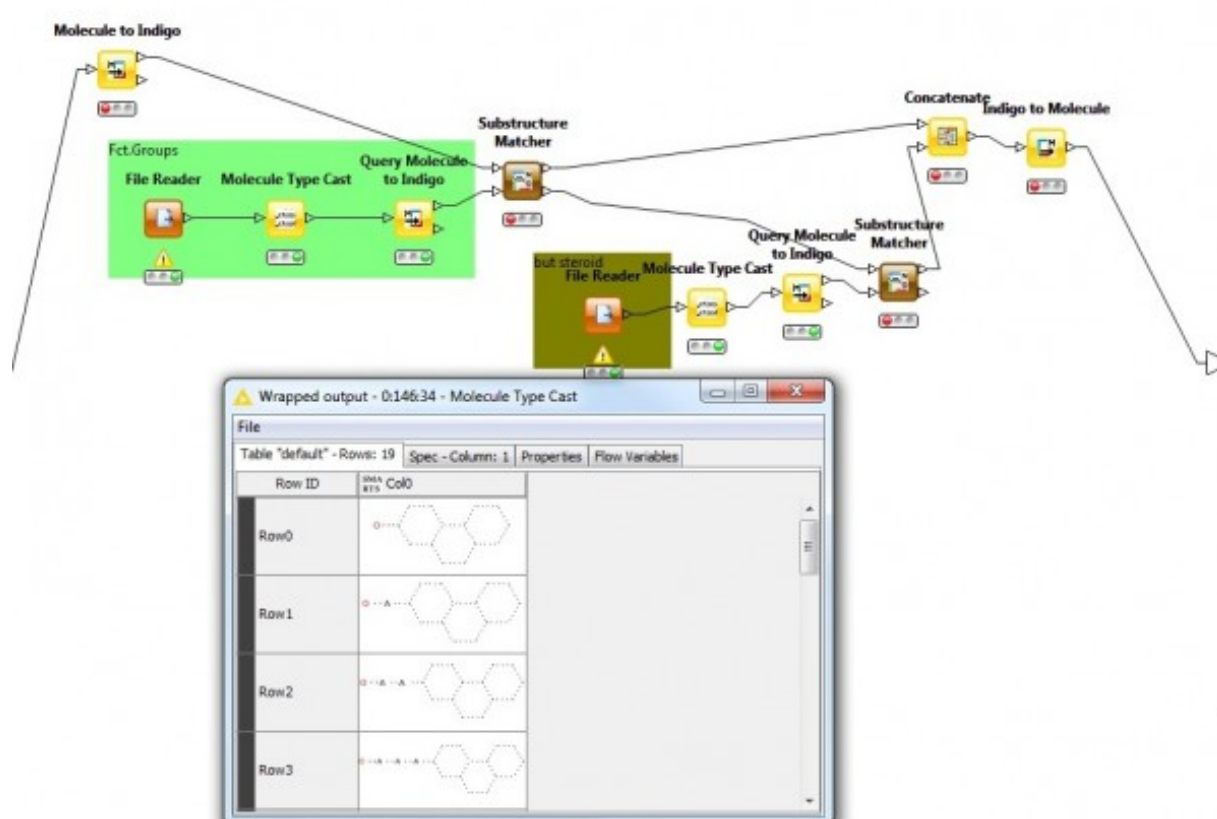


Figure 2: Substructure checking within the workflow.

Procedure Description

Biological Endpoint and Measurement:

This method is an in silico model, which provides information on the following, biologically relevant, endpoints:

- Receptor binding: Aryl hydrocarbon receptor (AHR)
- Receptor binding: Estrogen receptor (ER)
- Receptor binding: Farnesoid X receptor (FXR)
- Receptor binding: Glucocorticoid receptor (GR)
- Receptor binding: Liver X receptor (LXR)
- Receptor binding: Peroxisome proliferator-activated receptor (PPAR)
- Receptor binding: Pregnane X receptor (PXR)
- Receptor binding: Progesterone Receptor (PR)
- Receptor binding: Retinoic acid receptor (RAR)/Retinoid X receptor (RXR)
- Receptor binding: Thyroid Receptor (THR)
- Receptor binding: Vitamin D Receptor (VDR)

Endpoint Value:

Qualitative: Structural alert

Experimental System:

A computational model that predicts potential nuclear receptor binding and flags potential ligands, with additional alerts for NRs associated with the onset of hepatosteatosis.

3. Method Performance

Robustness of the Method

General performance measures:

All chemicals which have been identified as NR binders within ChEMBL were also successfully predicted by the model as ligands to their associated NRs, showing that the workflow was accurate at identifying structural alerts (Mellor et al., 2016).

4. Discussion

The ability to predict organ-level toxicity is becoming increasingly important to the long term goal of replacing animal use in determining a Lowest Observed (Adverse) Effect Level (LO(A)EL). The toxicity induced by nuclear receptor activation has been linked to many toxicities such as endocrine disruption and hepatic steatosis. Therefore, the links between the chemical structure of NR ligands and the biological responses of the NR they bind to have been elucidated, this has provided knowledge of the toxicological pathways linked to these NR, specifically to the adverse outcome of hepatic steatosis.

The workflow developed can be used to group chemicals, e.g. considering chemicals potentially linked to hepatic steatosis, for read-across and for hazard identification as part of a risk assessment process.

Nota bene: There are also other nodes to calculate descriptors and to alter and filter (sub)structures within KNIME (for example RDKit nodes).

5. Contact Details

Prof. Mark Cronin

Professor of Predictive Toxicology
School of Pharmacy and Biomolecular Sciences
Liverpool John Moores University
Corresponding author
Byrom Street
Liverpool L3 3AF
email: m.t.cronin@ljmu.ac.uk
telephone: + 44 151 231 2402
fax: + 44 151 231 2170
last contact update: 11 May 2016

Dr Fabian Steinmetz

School of Pharmacy and Biomolecular Sciences
Liverpool John Moores University
Method developer / user
Byrom Street
Liverpool L3 3AF

Dr Claire Mellor

School of Pharmacy and Biomolecular Sciences
Liverpool John Moores University
Method developer / user
Byrom Street

Liverpool L3 3AF

Ver n°	Date	Name	Type of change
1	12 November 2015	Prof. Mark Cronin	Data sheet created

6. Abbreviations & Definitions

AHR: Aryl Hydrocarbon Receptor

AOP: Adverse Outcome Pathway

AR: Androgen Receptor

ER: Estrogen Receptor

FXR: Farnesoid X Receptor

GR: Glucocorticoid Receptor

LO(A)EL: Lowest Observed (Adverse) Effect Level

LXR: Liver X Receptor

MW: Molecular weight

PPAR: Peroxisome Proliferator-Activated Receptor

PR: Progesterone Receptor

PXR: Pregnane X Receptor

RAR: Retinoic Acid Receptor

RB: Rotational bonds

RXR: Retinoid X receptor

NR: Nuclear Receptor

SA: Structural Alerts

KNIME: KoNstanz Information MinEr

PDB: Protein Data Bank

THR: Thyroid receptor

VAIM: Vertex adjacency information magnitude

VDR: Vitamin D Receptor

7. Bibliography

Method development:

- Bento, A.P., Gaulton, A., Hersey, A., Bellis, L.J., Chambers, J., Davies, M., Krüger, F.A., Light, Y., Mak, L., McGlinchey, S., Nowotka, M., Papadatos, G., Santos, R., Overington, J.P. (2013)
The ChEMBL bioactivity database: an update.
Nucleic Acid Research 42, 1-8
- ChEMBL (2014)
ChEMBL - a database of bioactive drug-like small molecules - <https://www.ebi.ac.uk/chembl/> (accessed 01.10.2014)
- Mellor, C.L., Steinmetz, F.P., Cronin M.T.D. (2015)
The Identification of Nuclear Receptors Associated with Hepatic Steatosis to Develop and Extend Adverse Outcome Pathways.
Critical Reviews in Toxicology 46, 138-152

- Mellor, C.L., Steinmetz, F.P., Cronin, M.T.D. (2016)
Using Molecular Initiating Events to Develop a Structural Alert Based Screening Workflow for Nuclear Receptor Ligands Associated with Hepatic Steatosis.
Chemical Research in Toxicology 29, 203-212
- Moya M., Gomez-Lechon M.J., Castell J.V., Jover R. (2010)
Enhanced steatosis by nuclear receptor ligands: A study in cultured human hepatocytes and hepatoma cells with a characterized nuclear receptor expression profile.
Chemico-Biological Interactions 184, 376-387
- Protein Data Bank (PDB) (2016)
RCSB Protein Data Bank - www.rcsb.org (access 21.07.2016)
- PyMOL (2016)
PyMOL: A molecular visualization system on an open source foundation, maintained and distributed by Schrödinger - <https://www.pymol.org/> (access 21.07.2016)
- Steinmetz F.P., Mellor C.L., Meini T., Cronin M.T.D. (2015)
Screening chemicals for receptor-mediated toxicological and pharmacological endpoints: Using public data to build screening tools within a KNIME workflow.
Molecular Informatics 34, 171-178

Last update: 10 June 2016

DB-ALM Method summary n° 178 : *In silico* profiler for covalent DNA binding

Genotoxicity/Mutagenicity

The *in silico* profiler for covalent DNA binding compiles mechanistic organic chemistry fragments (in the form of structural alerts) for the binding of organic compounds to DNA. This profiler is envisaged to be used to develop chemical categories based upon similar mechanisms as part of the adverse outcome pathway paradigm. Additionally, the profiler could be utilised in screening large data sets in order to identify chemicals with the potential to bind to DNA.

1. General Information

Corresponding author(s):

Dr Steve Enoch, Liverpool John Moores University

Date of data sheet creation: 12 November 2015

Last version number: 1

Last update: 12 November 2015

Proprietary and/or Confidentiality Issues:

None. The model is publicly available.

Status

Development

Development completed. Work carried out at Liverpool John Moores University with support from the European Chemicals Agency (ECHA) Service Contract No. ECHA/2008/20/ECA.203 and the FP7 COSMOS Project (Integrated *In Silico* Models for the Prediction of Human Repeated Dose Toxicity of Cosmetics to Optimise Safety), within the SEURAT 1 (Safety Evaluation Ultimately Replacing Animal Testing 1) cluster (2011-2015). More information on the COSMOS project is available on its dedicated website:

<http://www.cosmostox.eu>



Most of the structural alerts and associated mechanistic chemistry have been incorporated into the Organisation for Economic Co-operation and Development (OECD) (Q)SAR Application Toolbox, a software tool designed to fill data gaps in a regulatory environment without the need for further animal testing: <http://www.qsartoolbox.org>

The *in silico* profiler for covalent DNA binding has a number of structural alerts coded into SMARTS patterns and translated into KNIME (<http://www.knime.org/>) automated workflows as part of the COSMOS project. The KNIME workflow is available via COSMOS Space:

<http://cosmospace.cosmostox.eu>

A web version of the model was also implemented, allowing to execute the model via a web browser. It is available through the COSMOS KNIME WebPortal:

<http://knimewebportal.cosmostox.eu>

2. Method Definition

Purpose of the Method

TYPE OF TESTING	: Screening, Part of an integrated testing strategy
LEVEL OF TOXICITY ASSESSMENT	: Toxic potential

Context of use

The main goal of the *in silico* profiler is grouping chemicals into mechanism based categories centred on the knowledge of a number of key molecular initiating events (MIEs) for mutagenicity and genotoxic carcinogenicity as part of the adverse outcome pathway (AOP) paradigm (e.g. AOP 15 "Alkylation of DNA in male pre-meiotic germ cells leading to heritable mutations"; Yauk et al, 2016). Such categories allow for read-across and data gap-filling to be applied. The premise behind these structurally developed categories is that similar chemicals should have similar biological activities and therefore should have the same MIE.

Furthermore, the categories produced using *in silico* profilers can be supported by, and - screening large data sets in order to identify chemicals with the potential to bind to DNA - used to prioritise additional testing using *in vitro* and/or *in chemico* methods, within an integrated testing strategy (ITS) or an integrated approach to testing and assessment (IATA).

Such strategies can be used in hazard identification and risk assessment for regulatory purposes under REACH (EU, 2006) or the Cosmetics Directive (EU, 2009). To facilitate the use in this context the covalent DNA binding profiler was incorporated into *in silico* software tools such as the OECD QSAR Toolbox.

Scientific Principle of the Method

For both mutagenicity and carcinogenicity one of the fundamental steps is the formation of a covalent bond between nucleophile and electrophile (i.e. the biological nucleophilic DNA molecule and an exogenous chemical). DNA binding can thus be a molecular initiating event in an adverse outcome pathway leading to genotoxicity (e.g. AOP 15: *Alkylation of DNA in male pre-meiotic germ cells leading to heritable mutations*; Yauk et al, 2016).

Within the AOP paradigm, *in silico* methods contribute to identifying the key features of a chemical (structural alerts) that induce a MIE. A collection of structural alerts that induce the same MIE are considered to be an *in silico* profiler.

There are two types of profilers: mechanism-based profilers, as described here, and chemistry-based profilers. The first type is comprised of structural alerts relating to a MIE that are supported by experimental data illustrating how the alerts initiate the MIE, as for example also the *in silico* profilers for covalent protein binding (DB-ALM Method Summary No. 181) or mitochondrial toxicity (DB-ALM Method Summary No. 180). In comparison, the chemistry-based profiler, as for example the *in silico* profilers for potential nuclear receptor ligands and alerts towards hepatosteatosis (DB-ALM Method Summary No. 177) or hepatotoxicity (DB-ALM Method Summary No. 179), consists of structural alerts that have been associated with inducing toxicity; thus, in contrast to mechanism-based profilers, a chemistry-based profiler does not contain mechanistic information relating to how the observed toxicity is initiated.

Due to these intrinsic differences, the two types of profilers should be used for different purposes; mechanistic profilers should be used for category formation and read-across, whilst chemistry-based profilers should be used to screen an inventory to prioritise chemicals to undergo further (non-animal) testing. A category developed utilising a mechanism-based profiler enables missing toxicological data for a chemical to be filled using available information from analogues within the same category via read-across predictions (Enoch et al., 2011).

Model Description

The *in silico* profiler for covalent DNA binding was developed based on the review of current scientific knowledge relating to mutagenicity and genotoxic carcinogenicity structural alerts and mapping the existing structural alerts in terms of their relationships with mechanistic organic chemistry, i.e. identifying alerts from the published compilations related to covalent DNA binding. The mapping was performed to achieve maximum overlap and usability whilst restricting redundancy in the alerts, and to ensure that the alerts related to the molecular initiating event of covalent DNA binding.

The existing literature compilations of structural alerts for DNA binding (Mekenyan et al 2004; Mekenyan et al 2007; Serafimova et al 2007; Benigni et al 2008; Kazius et al 2005 and Kalgutkar et al 2005) have been used in the analysis to add the mechanistic organic chemistry associated with each of the structural alerts. The alerts cross six broad organic chemistry mechanisms (domains): acylation (AC), Michael addition (MA), Schiff base formation (SB), unimolecular aliphatic nucleophilic substitution (S_N1), bimolecular aliphatic nucleophilic substitution (S_N2) and radical mechanism (RD).

The mechanistic information is intended to outline how the alert can act as a direct electrophile or how it can be converted into an electrophile. Therefore, the important consideration within the mechanistic chemistry framework is the inclusion of potential metabolic activation.

A total of 111 structural alerts, supported by mechanistic information and meta data, have been created.

The identified mechanistic structural alerts were further defined as SMARTS (SMiles ARbitrary Target Specification, based on SMILES Simplified Molecular Input Line Entry System) patterns. SMARTS pattern allows for a definition of a sub-structure within a molecule, which is then used to check whether a chemical structure contains any of the pre-defined fragments.

The *in silico* profiler was implemented as KNIME workflow, and additionally as web version, allowing to execute the model via a web browser. It is available through the COSMOS KNIME WebPortal:

<http://knimewebportal.cosmostox.eu>

Table 1. Summary of the 111 structural alerts defined for covalent DNA binding.

AC - Acylation, MA - Michael addition, SB - Schiff base formation, S_N1 - nucleophilic substitution 1, S_N2 - nucleophilic substitution 2, RD - Radical mechanistic domain

The full description of the DNA binding profiler is available in Enoch et al. (2010 and 2012).

Structural alerts	SMARTS	Mechanism
Acyl-halide AC	<chem>[CH,\$(C(=O)[#6]))(=O)[F,Cl,Br,I]</chem>	Direct addition of an acyl halide
Alkyl-carbamyl halides AC	<chem>[NH2,\$([NH](C=O)[#6]),\$(N(C=O)([#6])[#6]))C(=O)[F,Cl,Br,I]</chem>	Direct addition of an acyl halide
Isocyanates AC	<chem>[NH,\$(N(=C=O)[#6]))=C=O</chem>	Isocyanates and isothiocyanates
Isothiocyanates AC	<chem>[NH,\$(N(=C=[SX1])[#6]))=C=[SX1]</chem>	Isocyanates and isothiocyanates
Thiazolidinediones AC	<chem>C1C(=O)[NH]C(=O)S1</chem>	P450 mediated activation to isocyanates or isothiocyanates
Formamides AC	<chem>[NH2,\$([NH]([CH]=O)[#6]),\$(N([CH]=O)([#6])[#6]))[CH](=O)</chem>	P450 mediated activation to isocyanates or isothiocyanates
Sulfonylureas AC	<chem>[#6]S(=O)(=O)[NH]C(=O)[NH2,\$([NH](C=O)[#6]),\$(N(C=O)([#6])[#6]))]</chem>	P450 mediated activation to isocyanates or isothiocyanates
Thioureas AC	<chem>c[NH]C(=O)[NH]c</chem>	P450 mediated activation to isocyanates or isothiocyanates
Benzylamines AC	<chem>c1ccccc1[CH4][NH2]</chem>	P450 mediated activation to isocyanates or isothiocyanates
1,1-Dihaloalkanes AC	<chem>[CX4][CH]([F,Cl,Br,I])[F,Cl,Br,I]</chem>	P450 mediated activation to isocyanates or isothiocyanates
α,β-Unsaturated aldehydes MA	<chem>[\$([CH2]),\$([CH]C)]=[CH,\$(C[CX4]))][CH]=O</chem>	Polarised alkenes

α,β-Unsaturated ketones MA	<chem>[\$([CH2]),\$([CH]C)=[CH,\$(C[CH4])]C(=O)[#6]);!\$([CH]1=CC(=O)C=CC1(=O));!\$([CH]1=CC(=O)C(=O)C=C1)]</chem>	Polarised alkenes
α,β-Unsaturated esters MA	<chem>[\$([CH2]),\$([CH]C)=[CH,\$(C[CH4])]C(=O)O[#6]</chem>	Polarised alkenes
α,β-Unsaturated amides MA	<chem>[\$([CH2]),\$([CH]C)=[CH,\$(C[CH4])]C(=O)[NH2,\$([NH](C=O)[#6]),\$(N(C=O)([#6])[#6])]</chem>	Polarised alkenes
Azocarbonamides MA	<chem>[NH2,\$([NH](C=O)[#6]),\$(N(C=O)([#6])[#6])]C(=O)N=NC(=O)[NH2,\$([NH](C=O)[#6]),\$(N(C=O)([#6])[#6])]</chem>	Polarised azo compounds
Quinones MA	<chem>[\$(C1(=O)[#6]=,:[#6]C(=O)[CH]=C1),\$(C1(=O)C(=O)[#6]=,:[#6][CH]=C1)]</chem>	Quinone and quinone type chemicals
Quinone methides MA	<chem>[\$(C1=CC(=O)[#6]=,:[#6]C1(=[CH][CX4])),\$(C1(=[CH][CX4])C(=O)[#6]=,:[#6]C=C1)]</chem>	Quinone and quinone type chemicals
Quinone imines MA	<chem>[\$(C1(=O)[#6]=,:[#6]\$(C=[NH]),\$(C=N[CX4]))[CH]=C1),\$(C1(=O)[\$(C=[NH]),\$(C=N[CX4]))[#6]=,:[#6][CH]=C1),\$(C1(=O)[\$(C=[NH]),\$(C=N[CX4])]C=[CH][#6]=,:[#6]1)]</chem>	Quinone and quinone type chemicals
Quinone diimines MA	<chem>[\$([CH]1=C\$(C=[NH]),\$(C=N[CX4]))[#6]=,:[#6]\$(C=[NH]),\$(C=N[CX4]))1),\$([CH]1=C\$(C=[NH]),\$(C=N[CX4]))[\$(C=[NH]),\$(C=N[CX4]))[#6]=,:[#6]1)]</chem>	Quinone and quinone type chemicals
Hydroquinones MA	<chem>[\$(c1([OH,\$(O[CH3]))c[CH]c([OH,\$(O[CH3]))])cc1),\$(c1([OH,\$(O[CH3]))c([OH,\$(O[CH3]))])cc[CH]c1)]</chem>	P450 mediated activation to quinones and quinone type chemicals
Alkyl phenols MA	<chem>[\$(c1([OH])ccc([CH3,\$([CH2][CX4]))])cc1),\$(c1([OH])c([CH3,\$([CH2][CX4])))cccc1)]</chem>	P450 mediated activation to quinones and quinone type chemicals
Methylenedioxyphenyl MA	<chem>[cH]1ccc(OCO2)c2c1</chem>	P450 mediated activation to quinones and quinone type chemicals
Arenes MA	<chem>[cH,\$(c(c)(c)[CX4])]1[cH][cH][cH,\$(c(c)(c)[CX4])[cH][cH,\$(c(c)(c)[CX4])]1</chem>	P450 mediated activation to quinones and quinone type chemicals
PAHs MA	<chem>[cH]1c[cH]c2cc3ccccc3cc2[cH]1</chem>	P450 mediated activation to quinones and quinone type chemicals
5-Alkoxyindoles MA	<chem>[cH]1cc([OH,\$(O[CH3]))cc2ccnc12</chem>	P450 mediated activation to quinones and quinone type chemicals
3-Methylindoles MA	<chem>[CH3]c1c[nH]c2ccccc12</chem>	P450 mediated activation to quinones and quinone type chemicals

Furans MA	[cH]1[cH,\$(c(c)(o)C)]o[cH,\$(c(c)(o)C)][cH,\$(c(c)(c)C)]1	
Thiophenes MA	[cH]1[cH,\$(c(c)(o)C)]s[cH,\$(c(c)(o)C)][cH,\$(c(c)(c)C)]1	P450 mediated activation of heterocyclic ring systems
Hydrazine S _N 1	[NH][NH2]	Carbenium ion formation
Aliphatic N nitro S _N 1	[CH3,\$([CH2]([NX3])[#6]),\$([CH]([NX3])([#6])[#6])][NH,\$(N([CX4])([NX3])[#6])[N+](=O)[O-]]	Carbenium ion formation
Triazines S _N 1	[NH2,\$([NH](N=N)[#6]),\$(N(N=N)([#6])[#6])N=N[CX4]	Carbenium ion formation
Diazoalkanes S _N 1	[\$([C-][N+]#N),\$(C=N=N)]	Carbenium ion formation
PAHs S _N 1	[cH]1[cH][cH]c2c(ccc3ccccc23)[cH]1	Carbenium ion formation
Nitroso S _N 1	[CH3,\$([CH2]([NX3])[#6]),\$([CH]([NX3])([#6])[#6])][NH,\$(N([CX4])([NX3])[#6])[NX2]=O]	Carbenium ion formation
Allyl benzenes S _N 1	C=C[CH2]c1ccccc1	Carbenium ion formation
Pyrroloizidine alkaloids S _N 1	[\$([CX4]C(=O)OCc1ccn2CCC(OC([CX4])=O)c12),\$([CX4]C(=O)OCC1=CCN2CCC(OC([CX4])=O)C12),\$([CX4]C(=O)OC\ C1=C\ CN([CH3])CCC(OC([CX4])=O)C1=O)]	Carbenium ion formation
a Halo ethers S _N 1	[#6][O,S][CH2,\$([CH]([O,S])([F,Cl,Br,I])[CX4]),\$(C([O,S])([F,Cl,Br,I])([CX4])[CX4])[F,Cl,Br,I]]	Carbenium ion formation
Primary aromatic amines S _N 1	[\$([c][NH2]);r6;\$([c]([a]C)([a]C)[NH2]);!\$([c]([c]C(=O)[OH])[NH2]);!\$([c]([c]S(=O)(=O)[OH])[NH2]);!\$([c]([a][c]S(=O)(=O)[OH])[NH2]);!\$([c]([a][a][c]S(=O)(=O)[OH])[NH2])]	Nitrenium ion formation
Secondary aromatic amines S _N 1	[\$([c][NH][#6]);r6;\$([c]([a]C)([a]C)[NH][#6]);!\$([c]([c]C(=O)[OH])[NH][#6]);!\$([c]([c]S(=O)(=O)[OH])[NH][#6]);!\$([c]([a][c]S(=O)(=O)[OH])[NH][#6]);!\$([c]([a][a][c]S(=O)(=O)[OH])[NH][#6])]	Nitrenium ion formation
Tertiary aromatic amines S _N 1	[\$([c]N([CH3])[CH3]);r6;\$([c]([a]C)([a]C)N([CH3])[CH3]);!\$([c]([c]C(=O)[OH])N([CH3])[CH3]);!\$([c]([c]S(=O)(=O)[OH])N([CH3])[CH3]);!\$([c]([a][c]S(=O)(=O)[OH])N([CH3])[CH3]);!\$([c]([a][a][c]S(=O)(=O)[OH])N([CH3])[CH3])]	Nitrenium ion formation
Aromatic nitros S _N 1	[\$([c][N+](=O)[O-]);r6;\$([c]([a]C)([a]C)[N+](=O)[O-]);!\$([c]([c]C(=O)[OH])[N+](=O)[O-]);!\$([c]([c]S(=O)(=O)[OH])[N+](=O)[O-]);!\$([c]([a][c]S(=O)(=O)[OH])[N+](=O)[O-]);!\$([c]([a][a][c]S(=O)(=O)[OH])[N+](=O)[O-])]	Nitrenium ion formation
Aromatic nitrosos S _N 1	[\$([c][NX2]=O);r6;\$([c]([a]C)([a]C)[NX2]=O);!\$([c]([c]C(=O)[OH])[NX2]=O);!\$([c]([c]S(=O)(=O)[OH])[NX2]=O);!\$([c]([a][c]S(=O)(=O)[OH])[NX2]=O);!\$([c]([a][a][c]S(=O)(=O)[OH])[NX2]=O)]	Nitrenium ion formation
Aromatic N hydroxylamines S _N 1	[\$(c[NH][OH]);r6]	Nitrenium ion formation

Aromatic azos S _N 1	<chem>[\$([c]N=[NH,\$(N(=N)[#6])]);r6;!\$([c]([a]C)([a]C)N=[NH,\$(N(=N)[#6])]);!\$([c]([c]C(=O)[OH])N=[NH,\$(N(=N)[#6])]);!\$([c]([c]S(=O)(=O)[OH])N=[NH,\$(N(=N)[#6])]);!\$([c]([a][c]S(=O)(=O)[OH])N=[NH,\$(N(=N)[#6])]);!\$([c]([a][a][c]S(=O)(=O)[OH])N=[NH,\$(N(=N)[#6])]);]</chem>	Nitrenium ion formation
Aromatic phenylureas S _N 1	<chem>[\$([c][NH]C(=O)[NH2,\$([NH](C=O)[#6]),\$(N(C=O)([#6])[#6])]);r6;!\$([c]([a]C)([a]C)[NH]C(=O)[NH2,\$([NH](C=O)[#6]),\$(N(C=O)([#6])[#6])]);!\$([c]([c]C(=O)[OH])[NH]C(=O)[NH2,\$([NH](C=O)[#6]),\$(N(C=O)([#6])[#6])]);!\$([c]([c]S(=O)(=O)[OH])[NH]C(=O)[NH2,\$([NH](C=O)[#6]),\$(N(C=O)([#6])[#6])]);!\$([c]([a][c]S(=O)(=O)[OH])[NH]C(=O)[NH2,\$([NH](C=O)[#6]),\$(N(C=O)([#6])[#6])]);!\$([c]([a][a][c]S(=O)(=O)[OH])[NH]C(=O)[NH2,\$([NH](C=O)[#6]),\$(N(C=O)([#6])[#6])]);]</chem>	Nitrenium ion formation
Aromatic ester hydroxylamines S _N 1	<chem>[\$([c][NH]OC(=O)[#6]);r6;!\$([c]([a]C)([a]C)[NH]OC(=O)[#6]);!\$([c]([c]C(=O)[OH])[NH]OC(=O)[#6]);!\$([c]([c]S(=O)(=O)[OH])[NH]OC(=O)[#6]);!\$([c]([a][c]S(=O)(=O)[OH])[NH]OC(=O)[#6]);!\$([c]([a][a][c]S(=O)(=O)[OH])[NH]OC(=O)[#6]);]</chem>	Nitrenium ion formation
Primary heterocyclic amines S _N 1	<chem>[\$(c[NH2]);r5]</chem>	Nitrenium ion formation
Secondary heterocyclic amines S _N 1	<chem>[\$(c[NH][#6]);r5]</chem>	Nitrenium ion formation
Tertiary heterocyclic amines S _N 1	<chem>[\$(cN([CH3])[CH3]);r5]</chem>	Nitrenium ion formation
Heterocyclic nitro S _N 1	<chem>[\$(c[N+](=O)[O-]);r5]</chem>	Nitrenium ion formation
Heterocyclic nitroso S _N 1	<chem>[\$(c[NX2]=O);r5]</chem>	Nitrenium ion formation
Heterocyclic N-hydroxylamines S _N 1	<chem>[\$(c[NH][OH]);r5]</chem>	Nitrenium ion formation
Heterocyclic azos S _N 1	<chem>[\$(cN=[NH,\$(N(=N)[#6])]);r5]</chem>	Nitrenium ion formation
Heterocyclic phenylureas S _N 1	<chem>[\$(c[NH]C(=O)[NH2,\$([NH](C=O)[#6]),\$(N(C=O)([#6])[#6])]);r5]</chem>	Nitrenium ion formation
Heterocyclic ester hydroxylamine S _N 1	<chem>[\$(c[NH]OC(=O)[#6]);r5]</chem>	Nitrenium ion formation
Nitroso S _N 1	<chem>N(c)(c)[NX2]=O</chem>	Nitrosation
Aliphatic tertiary amines S _N 1	<chem>[NX3](C)(C)C</chem>	Iminium ion formation
Thiophenes S _N 2	<chem>[cH,\$(c(c)(c)C)]1[cH,\$(c(c)(c)C)]s[cH][cH]1</chem>	P450 mediated epoxidation
Coumarins S _N 2	<chem>[cH]1[cH]c(=O)oc2ccccc12</chem>	P450 mediated epoxidation
Mustards S _N 2	<chem>[NX3,SX2][CH2,\$([CH]([NX3,SX2])([CX4])[CX4])][CH2,\$([CH]([CX4])([Cl,Br,I])[CX4])][Cl,Br,I]</chem>	Episulfonium ion formation
1,2-Dihaloalkanes S _N 2	<chem>[Cl,Br,I][CH2,\$([CH]([Cl,Br,I])([CX4])[CX4])][CH2,\$([CH]([Cl,Br,I])([CX4])[CX4])][Cl,Br,I]</chem>	Episulfonium ion formation
Alkyl nitrate S _N 2	<chem>[CH3,\$([CH2](O)[#6]),\$([CH](O)([#6])[#6])][O][NX2]=O</chem>	S _N 2 at an sp ³ carbon atom
Alkyl carbamates S _N 2	<chem>[CH3,\$([CH2]([NX3])[CX4]),\$([CH]([NX3])([CX4])[CX4])][NX3]([CH3])C(=O,SX2)[O,SX2][CX4]</chem>	S _N 2 at an sp ³ carbon atom

Aliphatic halide S _N 2	[CH3,\$([CH2]([F,Cl,Br,I])[#6]),\$([CH]([F,Br,Cl,I])([#6])[#6]);!\$([CH2,CH]C=O);!\$([CH2,CH][CX4][NX3,SX2]);!\$([CH2,CH][CX4][F,Cl,Br,I])[F,Cl,Br,I]	S _N 2 at an sp3 carbon atom
Sulfonates S _N 2	[#6]S(=O)(=O)O[CH3,\$([CH2](O)[#6]),\$([CH](O)([#6])[#6])]	S _N 2 at an sp3 carbon atom
Sulfates S _N 2	[#6]OS(=O)(=O)O[CH3,\$([CH2](O)[#6]),\$([CH](O)([#6])[#6])]	S _N 2 at an sp3 carbon atom
Phosphates S _N 2	[#6][O,SX2]P(=[O,SX1])([O,SX2][#6])[O,SX2][CH3,\$([CH2](O)[#6]),\$([CH](O)([#6])[#6])]	S _N 2 at an sp3 carbon atom
Phosphonates S _N 2	[#6][O,SX2]P(=[O,SX1])([#6])[O,SX2][CH3,\$([CH2](O)[#6]),\$([CH](O)([#6])[#6])]	S _N 2 at an sp3 carbon atom
Phosphonic esters S _N 2	[#6][PX4](=O)O[CH3,\$([CH2](O)[#6]),\$([CH](O)([#6])[#6])]	S _N 2 at an sp3 carbon atom
Lactones S _N 2	C1CC(=O)O1	Ring opening S _N 2
Sultones S _N 2	[\$(C1OS(=O)(=O)CC1),\$(C1OS(=O)(=O)CCC1)]	Ring opening S _N 2
Nitrosos S _N 2	[CX4][NX3]([#6])[NX2]=O	Nitrosation
Epoxides S _N 2	C1OC1	Direct acting epoxides and related
Aziridines S _N 2	C1[NX3]C1	Direct acting epoxides and related
Sulfaranes S _N 2	C1[SX2]C1	Direct acting epoxides and related
Halogenated polarised alkenes S _N 2	[CH2,\$([CH](=[CH])[#6,F,Cl,Br,I])=[CH][F,Cl,Br,I,\$(O(C=C)[a])]	Epoxidation of aliphatic alkenes
N-acyloxy-N-alkoxyamides S _N 2	[a]C(=O)N(O[#6])OC(=O)[#6]	S _N 2 at a nitrogen atom
Thioureas S _N 2	[#6][NH]C(=S)[NH][#6]	P450 mediated sulfoxidation
Mono-aldehydes SB	[CH2,\$([CH](=O)[CX4])]=O	Direct acting Schiff base formers
a,b-Dicarbonyl SB	[CH,\$(C(=O)[#6]);!\$(C(=O)C=C)(=O)[CH,\$(C(=O)[#6]);!\$(C(=O)C=C)]=O	Direct acting Schiff base formers
Ethylenediamines SB	[\$([NH2,\$([NH]([CH2])[CX4]))[CH2][CH2][NH2]),\$([NH]1[CH2][CH2][NH][CH2][CH2]1)]	Chemicals activated by P450 to glyoxal
Ethanolamines SB	[\$([NH2,\$([NH]([CH2])[CX4]))[CH2][CH2][OH]),\$([NH]1[CH2][CH2]O[CH2][CH2]1)]	Chemicals activated by P450 to glyoxal
N-methylol derivatives SB	[NH2,\$([NH]([CH2])[#6]),\$(N([CH2])([#6])[#6])[CH2][OH]	Chemicals activated by P450 to mono aldehydes
Thiazoles SB	[cH]1ncs[cH]1	Chemicals activated by P450 to mono aldehydes

Benzylamines SB	<chem>c1ccccc1[CH2][NH2]</chem>	Chemicals activated by P450 to mono aldehydes
Acetates AC	<chem>[#6]C(=[O,SX2])[O,SX2]\$([a;r6]),\$([a;r5]),\$(C([O,SX2])=C),\$(C([O,SX2])#C)</chem>	Direct acylation
Acetohydroxamic acids and related AC	<chem>[#6,#7]C(=O)[NX3H,\$([NX3](C=O)([OH])[#6])][OH]</chem>	Activated to N hydroxy
Alkyl carbamates and related AC	<chem>[#6][OX2,NX3]C(=O)[NX3H2,\$([NX3H][CX4])]</chem>	Activated to N hydroxy
Aromatic and aliphatic diazoniums AC	<chem>\$([#6][NX2+]#NX1),\$([#6][NX2]=[NX2][OH])</chem>	Azo coupling
Ceracemides AC	<chem>[NH2,\$([NH](C=O)[#6]),\$(N(C=O)([#6])[#6])C(=O)O[NX3]([\$(CH)=O),\$(C([NX3])(=O)[#6]))C(=O)[NH2,\$([NH](C=O)[CH3]),\$(N(C=O)([CH3])[CH3])]</chem>	Activated to N hydroxy
Nitrones AC	<chem>[CH3,CH2,CH][CX3]=[NX2][OH]</chem>	Nitrones and related
Phosphinic halides esters AC	<chem>[PX4](=O)[F,Cl,Br,I,\$(Oa)]</chem>	Acylation at a P atom
Sulphonyl halides AC	<chem>[#6]S(=O)(=O)[F,Cl,Br,I]</chem>	Direct acylation
Cyclohexadiene 1,2 diols MA	<chem>C1([OH])[#6]=,[#6]C=CC1[OH]</chem>	Pre-quinones and quinone type chemicals
Napthalenes and related MA	<chem>[c;R1]1[c;R1][c;R2]2[a;R1][a;R1][a;R1][a;R1][c;R2]2[cH][cH]1</chem>	Pre-quinones and quinone type chemicals
Polarised alkene cyano MA	<chem>[CH2,CH]=CC#N</chem>	Polarised alkenes
Pyranones MA	<chem>\$([cH]1cc(=[O,NH])cc[o,n]1),\$([cH]1cc(=[O,NH])[o,n]cc1),\$(c1cc(=[O,NH])[o,n][cH]c1)</chem>	Quinones and quinone type chemicals
Quinolines MA	<chem>\$([cH]1[n;R1][c;R2]2[c;R1][c;R1][c;R1][c;R2]2[c;R1][cH]1),\$([c;R1]1[n;R1][c;R2]2[c;R1][c;R1][c;R1][c;R1][c;R2]2[cH][cH]1)</chem>	Pre quinones and quinone type chemicals
Quinone diimine MA	<chem>\$([CH]1=C\$(C=[NX2H]),\$(C=[NX2][#6,#7])[#6]=,[#6]\$(C=[NX2H]),\$(C=[NX2][#6,#7])1),\$([CH]1=C\$(C=[NX2H]),\$(C=[NX2][#6,#7])\$\$(C=[NX2H]),\$(C=[NX2][#6,#7])[#6]=,[#6]1))</chem>	Quinones and quinone type chemicals
Quinone dioximes MA	<chem>\$(C1(=[NX2][OH])[#6]=,[#6]C(=[NX2][OH])[CH]=C1),\$(C1(=[NX2][OH])C(=[NX2][OH])[#6]=,[#6][CH]=C1)</chem>	Quinones and quinone type chemicals
1,3-Dithianes SB	<chem>C1S[CH2,\$([CH][#6])]SCC1</chem>	Activated to di carbonyl
Aromatic and related triazenes S _N 1	<chem>[c;r6][NX2]=[NX2][NX3H2,\$([NX3H][#6]),\$([NX3]([CH3])[CH3])]</chem>	Nitrenium ion formation
N-Acetyl phenyl hydrazines S _N 1	<chem>c[NH][NH][CX3](=O)[CX4]</chem>	Carbenium ion formation
1,1-Dihalos S _N 2	<chem>[CH2,\$([CH]([F,Cl,Br,I])([F,Cl,Br,I])[#6])][F,Cl,Br,I][F,Cl,Br,I]</chem>	S _N 2 at an sp ³ carbon atom
Activated haloalkyls S _N 2	<chem>[CX4]([F,Cl,Br,I])\$([NX3+](=O)[O-]),\$([CX3]([CX4])(=O)[#6]),\$([CH]([CX4])=O),\$(a[CX4])</chem>	S _N 2 at an sp ³ carbon atom

Aliphatic nitros S _N 2	<chem>[\$([CH])([NX3+](=O)[O-])([#6])[#6]),\$(C([NX3+](=O)[O-])([#6])([#6])[#6])[NX3+](=O)[O-]</chem>	S _N 2 at an sp ³ carbon atom
Allyl acetates and related S _N 2	<chem>[#6]C(=[O,SX1])[O,SX2][CH2,\$([CH]([O,SX2])[#6])[c,CX3,CX2]</chem>	S _N 2 at an sp ³ carbon atom
Haloalkenes S _N 2	<chem>C=C[F,Cl,Br,I]</chem>	S _N 2 reaction at a sp ² carbon
N-halos S _N 2	<chem>[c;r6][NH][F,Cl,Br,I]</chem>	S _N 2 reaction at a halo atom
Peroxides RD	<chem>[\$([#6][OX2][OX2][#6]),\$([#6][OX2][OX2H]),\$([OX2H][OX2H])]</chem>	Radical
Quindioxin and related RD	<chem>[O-][n+]1cc[n+](O-)c2ccccc12</chem>	Radical

Procedure Description

Biological Endpoint and Measurement:

This method is an *in silico* model, which provides information on the following, biologically relevant, endpoints:

DNA damage: formation of covalent adducts

Endpoint Value:

Qualitative: Structural alert

Experimental System:

In silico profiler for covalent DNA binding

3. Method Performance

Robustness of the Method

General performance measures:

Enoch and Cronin reported in 2012 an extensive evaluation and new structural alerts for chemical category formation including covalent and non-covalent mechanisms of DNA binding.

A total of 2983 chemicals with positive Ames-test results from Kazius et al. (2005) were profiled with the first set of structural alerts for covalent DNA binding developed by Enoch and Cronin (2010). At least one electrophilic mechanism could be assigned to 2484 chemicals (83%).

A detailed mechanistic study of the remaining 498 chemicals with positive Ames-test results identified 26 new structural alerts for covalent-binding mechanisms, two for radical-mediated mechanisms and 12 structural alerts related to chemical classes of polycyclic aromatic hydrocarbons with common structural motifs that could explain mutagenicity by a non-covalent intercalation mechanism.

Only for 150 chemicals with positive Ames-test data (5%) no clear mechanism could be hypothesised from the available literature.

4. Discussion

The formation of a covalent bond between nucleophile (DNA) and electrophile (chemical) is one of the fundamental steps for both mutagenicity and carcinogenicity. The chemical adducts can lead to mutations, steric hindrance or strand breaks either directly by altering DNA sequence at the adduct site or causing failure of the attempted DNA repair (see e.g. Yauk C et al., 2016).

The mechanistic importance of this chemical reaction makes the mechanistic alert approach the natural choice for developing the profiler for covalent DNA binding. Moreover, the mechanistic domain overlap between corresponding structural alerts in the literature compilations has been thoroughly assessed, so that at least for the Ames test positive chemicals, only a very small minority could not be explained by the covalent or non-covalent interaction with DNA. The analysis by Enoch and Cronin (2012) ensured that for a given structural alert the maximum mechanistic information (and thus domain) was extracted.

The use of mechanistic chemistry based alerts enables transparent category formation. Moreover, the use of mechanistic classes (usually as defined by the presence of common reactive functional groups) can help to develop a number of (quantitative) structure- activity relationship ((Q)SAR) models for mutagenicity and carcinogenicity (see Passerini 2003; Benigni et al., 2008). These approaches play an important role in contributing towards reducing the use of animals in regulatory toxicology. Applying read-across to predict the toxicity of chemicals for which no toxicological data may exist, but might be available for other members of the same category, is already an accepted practice in REACH dossiers for genotoxicity and carcinogenicity endpoints. For more information consult the *"Analysis of carcinogenicity testing for regulatory purposes in the European Union"* (Madia et al., 2016).

5. Contact Details

Dr Steve Enoch

School of Pharmacy and Biomolecular Sciences
 Liverpool John Moores University
 Corresponding author, Method developer / user
 Byrom Street
 Liverpool L3 3AF
 email: s.j.enoch@ljmu.ac.uk
 telephone: +44(0)151 231 2164

Ver n°	Date	Name	Type of change
1	12 November 2015	Dr Steve Enoch	Data sheet created

6. Abbreviations & Definitions

AC Acylation

AOP Adverse Outcome Pathway

COSMOS EU Integrated Project: COSMOS developed an integrated suite of computational tools to predict the effects of long-term exposure to chemicals in humans based on *in silico* calculation and estimation techniques.

Website: <http://www.cosmostox.eu> .

ITS Integrated Testing Strategy

IATA Integrated Approach to Testing and Assessment

KNIME KoNstanz Information MinEr

MA Michael Addition

MIE Molecular Initiating Event

RD Radical Mechanism

SB Schiff Base formation

SEURAT Safety Evaluation Ultimately Replacing Animal Testing

SMARTS SMiles ARbitrary Target Specification

SMILES Simplified Molecular-Input Line-Entry System

S_N1 Unimolecular Aliphatic Nucleophilic Substitution

S_N2 Bimolecular Aliphatic Nucleophilic Substitution

7. Bibliography

Method development:

- Benigni R. and Bossa C. (2008)
Structure alerts for carcinogenicity, and the Salmonella assay system: A novel insight through the chemical relational databases technology
Mutation Research 659, 248-261
- Enoch S.J., Cronin M.T.D. (2012)
Development of new structural alerts suitable for chemical category formation for assigning covalent and non-covalent mechanisms relevant to DNA binding.
Mutation Research 743(1-2), 10-19
- Enoch S.J., Cronin M.T.D. (2010)
A review of the electrophilic reaction chemistry involved in covalent DNA binding.
Critical Reviews in Toxicology 40(8), 728-748
- Enoch S.J., Cronin M.T.D., Ellison C.M. (2011)
The use of a chemistry based profiler for covalent DNA binding in the development of chemical categories for read-across for genotoxicity.
Alternatives to Laboratory Animals (ATLA) 39, 131-145
- Kalgutkar A.S., Gardner I., Obach R.S., Shaffer C.L., Callegari E., Henne, K.R., Mutlib A.E., Dalvie D. K., Lee J.S., Nakai Y., O'Donnell J.P., Boer, J., Harriman S.P. (2005)
A comprehensive listing of bioactivation pathways of organic functional groups.
Current Drug Metabolism 6, 161-225
- Kazius J., McGuire R. and Bursi R. (2005)
Derivation and validation of toxicophores for mutagenicity prediction.
Journal of Medicinal Chemistry 48, 312-320
- Mekenyan O., Dimitrov S., Serafimova R., Thompson E., Kotov S., Dimitrova N. and Walker J.D. (2004)
Identification of the structural requirements for mutagenicity by incorporating molecular flexibility and metabolic activation of chemicals I: TA100 Model
Chemical Research in Toxicology 17, 753-766
- Mekenyan O., Todorov M., Serafimova R., Stoeva S., Aptula A., Finking R. and Jacob E. (2007)
Identifying the structural requirements for chromosomal aberration by incorporating molecular flexibility and metabolic activation of chemicals.
Chemical Research in Toxicology 20, 1927-1941
- Serafimova R., Todorov M., Pavolo T., Kotov S., Jacob E., Aptula A. and Mekenyan O. (2007)
Identification of the structural requirements for mutagenicity, by incorporating molecular flexibility and metabolic activation of chemicals. II. General Ames mutagenicity model.
Chemical Research in Toxicology 20, 662-676

General references:

- EU (2009)
Regulation (EC) No 1223/2009 of the European Parliament and of the Council of 30 November 2009 on cosmetic products. [Link to document](#) (last access 13.09.2016)
Official Journal of the European Communities

- EU (2006)
Regulation (EC) No 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH). [Link to document](#) (last access 16.09.2013)
Official Journal of the European Communities L396
- Madia, F., Worth, A., and Corvi, R. (2016)
Analysis of carcinogenicity testing for regulatory purposes in the European Union
[JRC publication repository](#)
- Passerini L. (2003)
QSARs for individual classes of chemical mutagens and carcinogens, in: R. Benigni (Ed.), Quantitative Structure–Activity Relationships (QSAR) Models of Mutagens and Carcinogens.
- Yauk C., Lambert I., Marchetti F., Douglas G. (2016)
AOP:15 - Alkylation of DNA in male pre-meiotic germ cells leading to heritable mutations. Access: https://aopwiki.org/wiki/index.php/File:Aop15-WNT_TFHA_Review.pdf

Last update: 15 November 2015

DB-ALM Method summary n° 179 : *In silico* profiler for hepatotoxicity

Systemic Toxicity

The *in silico* profiler for hepatotoxicity consists of 16 structural alerts coded into SMARTS patterns. The alerts can be used in the further screening of compounds to highlight potential hepatotoxicity. The profiler is envisaged to be used to develop chemical categories based upon similar mechanisms as part of the adverse outcome pathway paradigm.

1. General Information

Corresponding author(s):

Prof. Mark Cronin, Liverpool John Moores University

Date of data sheet creation: 12 November 2015

Last version number: 1

Last update: 12 November 2015

Proprietary and/or Confidentiality Issues:

None. The model is publicly available.

Status

Development

Development completed. Work carried out at Liverpool John Moores University with support from the FP7 COSMOS Project (Integrated *In Silico* Models for the Prediction of Human Repeated Dose Toxicity of Cosmetics to Optimise Safety), within the SEURAT-1 (Safety Evaluation Ultimately Replacing Animal Testing 1) cluster (2011-2015). More information on the COSMOS project is available on its dedicated website:

<http://www.cosmostox.eu>



The *in silico* profiler for hepatotoxicity has a number of structural alerts coded into SMARTS patterns and translated into KNIME (<http://www.knime.org/>) automated workflows as part of the COSMOS project. The KNIME workflow is available via COSMOS Space:

<http://cosmospace.cosmostox.eu>

A web version of the model was also implemented, allowing to execute the model via a web browser. It is available through the COSMOS KNIME WebPortal:

<http://knimewebportal.cosmostox.eu>

2. Method Definition

Purpose of the Method

TYPE OF TESTING	: Screening, Part of an integrated testing strategy
LEVEL OF TOXICITY ASSESSMENT	: Toxic potential

Context of use

Liver is exposed to the xenobiotics taken up via the oral exposure route or injection. Many undergo biotransformation in the liver and may cause hepatotoxicity before or after metabolism, which is a major concern for the safety of chemical and pharmaceutical products.

The main goal of the *in silico* profiler is grouping chemicals into mechanism based categories centred on the knowledge of a number of key molecular initiating events (MIEs) for hepatotoxicity as part of the

adverse outcome pathway (AOP) paradigm (e.g. AOP 38 "Protein alkylation leading to liver fibrosis"; Landesmann, 2012 and 2016). Such categories allow for read-across and data gap filling to be applied. The premise behind these structurally developed categories is that similar chemicals should have similar biological activities and therefore should have the same MIE.

Furthermore, the categories produced using *in silico* profilers can be supported by, and - screening large data sets in order to identify chemicals with potential hepatotoxicity - used to prioritise additional testing using *in vitro* and/or *in chemico* methods, within an integrated testing strategy (ITS) or an integrated approach to testing and assessment (IATA). Such strategies can be used in hazard identification and risk assessment for regulatory purposes such as REACH (EU, 2006) or the Cosmetics Directive (EU, 2009) and could be incorporated into *in silico* decision-supporting software tools such as the OECD QSAR Toolbox.

Scientific Principle of the Method

Within the AOP paradigm, *in silico* methods contribute to identifying the key features of a chemical (structural alerts) that induce a MIE. A collection of structural alerts that induce the same MIE are considered to be an *in silico* profiler.

There are two types of profilers: mechanism-based profilers and chemistry-based profilers. The first type is comprised of structural alerts relating to MIE that are supported by experimental data illustrating how the alerts initiate the MIE, for example the *in silico* profilers for covalent DNA binding (DB-ALM Method Summary No. 178), covalent protein binding (DB-ALM Method Summary No. 181) or mitochondrial toxicity (DB-ALM Method Summary No. 180).

In comparison, the chemistry-based profiler, as described here and for example the *in silico* profiler for potential nuclear receptor ligands and alerts towards hepatosteatosis (DB-ALM Method Summary No. 177) consists of structural alerts that have been associated with inducing toxicity; thus, in contrast to mechanism-based profilers, a chemistry-based profiler does not necessarily contain mechanistic information relating to how the observed toxicity is initiated.

Due to these intrinsic differences, the two types of profiler should be used for different purposes; mechanistic profilers should be used for category formation and read-across, whilst chemistry-based profilers should be used to screen an inventory to prioritise chemicals to undergo (non-animal) testing. A category developed utilising a mechanism-based profiler enables missing toxicological data for a chemical to be filled using available information from analogues within the same category via read-across predictions (Enoch et al 2011).

Model Description

The structural alerts were developed based on grouping chemicals from a hepatotoxicity dataset compiled by Fourches et al. (2010) into 16 categories, investigating the mechanism(s) by which these compounds cause hepatotoxicity and proposing a mechanistic rationale, where possible, to obtain mechanistically supported structural alerts for potential hepatotoxins.

This was achieved by grouping chemicals based upon their structural similarity, followed by a literature search to elucidate mechanistic information for the chemicals in categories associated with liver toxicity.

The identified structural alerts were defined as SMARTS (SMiles ARbitrary Target Specification, based on SMILES - Simplified Molecular-Input Line-Entry System) patterns. SMARTS pattern allows for a definition of a sub-structure within a molecule, which is then used to check whether a chemical structure contains any of the pre-defined fragments.

The *in silico* profiler was implemented as KNIME workflow, and additionally as web version, allowing to execute the model via a web browser. It is available through the COSMOS KNIME WebPortal:

<http://knimewebportal.cosmostox.eu>

Hepatotoxicity Dataset

One of the largest hepatotoxicity datasets available within the open literature was used to develop the structural alerts. This dataset was compiled by Fourches et al. (2010), who mined the open literature and got approximately 650,000 MEDLINE titles and abstracts containing one or more predetermined terms relating to drug-induced liver injury (DILI). A combination of lexical and linguistic tools was used by Fourches et al. to extract relationships between therapeutic compounds and a range of liver pathologies (including hepatitis and focal necrosis, hepatic physiological observations such as gluconeogenesis and cell growth). The data mining procedure considered data from *in vivo* and *in vitro* sources relating to observed hepatotoxicity. Following a data curation process (removal of inorganics, duplicates and

mixtures), a final dataset of 951 substances was obtained. Each of the 951 compounds is accompanied with data for three species groupings: human, rodent (rat, mouse, hamster, and guinea pig) and non-rodent (dog, cat, pig, monkey, etc.). To develop the *in silico* profiler described by Hewitt et al. (2013), only the human toxicity data were used for 650 compounds classified as positive for hepatotoxicity in humans by Fourches et al. (2010).

Strategy for category formation and structural alert development as defined by Hewitt et al. (2013)

Step 1: Acquire appropriate training data

The first step in any modelling process is to gather appropriate chemical and toxicological data. In terms of structural alert development this must consist of chemical structure and toxicological data relevant to the endpoint of interest. Then, all chemical structures were encoded into SMILES strings, neutralised, and salts were removed.

Step 2: Initial structural similarity-based category formation

The structural similarity of each compound to the remaining dataset members was calculated by means of an atom environment / Tanimoto coefficient approach using the freely available Toxmatch structural similarity tool (version 1.07) (Ideacon Ltd., 2012). The resulting similarity matrix was assessed manually and valid categories were highlighted if an active (hepatotoxic) query compound was matched with five or more compounds with a similarity index of 0.6 or greater (yielding categories of six or more compounds). The choice of similarity threshold was important as this directly influenced the level of structural restriction/diversity allowed within resulting categories.

Step 3: Manual validation of initial categories

Despite structural similarity methods often yield a significant number of categories, and a number of these categories contain members which are not truly similar to the query compound. In such instances, the observed structural diversity would likely lead to further diversity in the observed mechanisms of hepatotoxicity making such categories meaningless. In order to eliminate such categories from the analysis and ensure those remaining are valid, a simple three-tiered manual inspection process was applied:

1. Noisy categories containing unsuitable (structurally heterogeneous) category members were removed.
2. Duplicate categories (containing identical members but using different initiating compounds) were removed.
3. Finally, upon inspection, a small number of additional structural alerts were suggested based on the well-known hepatotoxicity of certain compound classes (e.g. the nitrogen mustards). These were not identified using structural similarity alone.

As a result of this process, 16 manually validated structural categories for human hepatotoxicity were developed.

Step 4: Elucidation of structural alerts

Following the generation of validated structural categories, the next step was to use these to develop structural alerts for hepatotoxicity. In order to do this, the common structural feature of each validated category was found via manual inspection of chemical structures and the corresponding structural alert was created in the form of a SMARTS notation (see **Table 1**).

Step 5: Using structural alerts to regenerate mechanistic categories

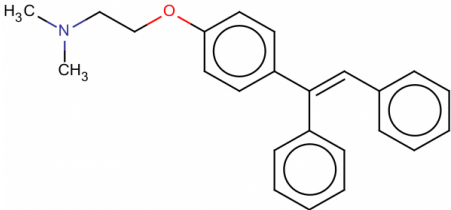
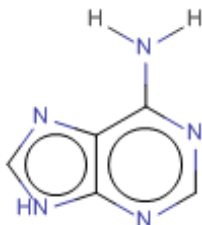
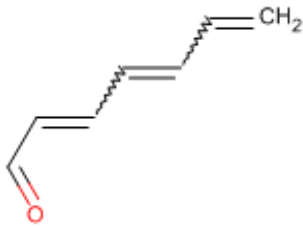
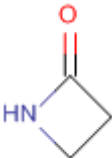
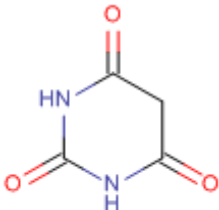
In this step, the original dataset was re-screened using the resulting structural alerts to re-populate each category. In this way, all compounds containing a given structural fragment will enter the corresponding category, showing the full scope of the proposed structural alert and maximising the utilisation of the available toxicological data. In some cases, this alert returns fewer compounds than were present in the initial category. This was a result of these categories being further restricted in terms of structural criteria. This stage could also be seen as an action to correct for setting a similarity threshold too high during the formation of initial categories. This could, therefore, allow for the use of higher levels of structural similarity without any resultant reduction in the number of category members.

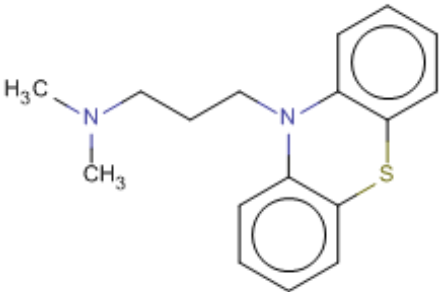
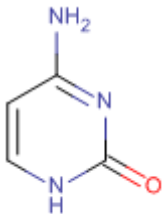
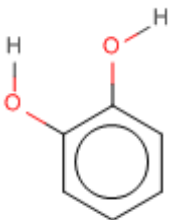
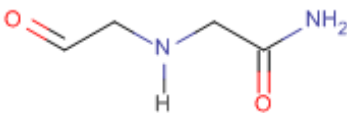
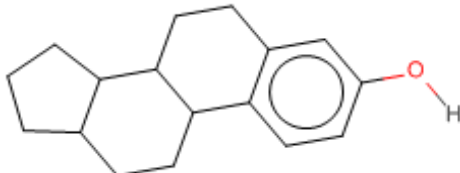
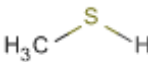
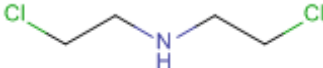
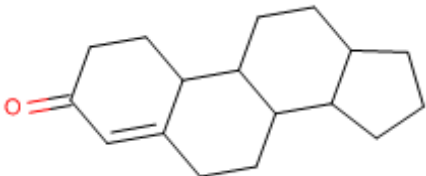
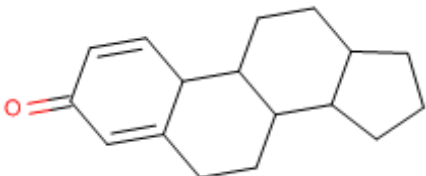
Step 6: Examine regenerated categories (alert verification) and propose mechanistic rationale

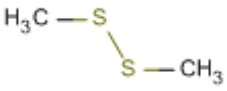
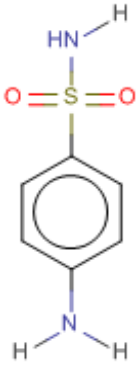
At this point, each of the regenerated categories was taken in turn and a detailed examination of all category members was undertaken in order to firstly verify the performance of each alert and, secondly, to derive a mechanistic hypothesis for the observed hepatotoxicity. This verification element is in place to ensure each alert was developed (coded) correctly and returns categories containing the structural feature of interest. This step is also beneficial for the identification and assessment of possible false positive and negative compounds. Of course, given the lack of current knowledge into the mechanisms of hepatotoxicity, and the often idiosyncratic nature of the effects, it was accepted from the outset that a mechanistic rationale may not be possible in all cases. Moreover, it is also probable that some hepatotoxic compounds may elicit their toxicities via multiple mechanisms, making this process tougher still.

The name, structure and SMARTS notification of each structural alert are presented in **Table 1** below. The full description of the hepatotoxicity profiler is available in Hewitt et al (2013).

Table 1. Summary of the 16 structural alerts identified for hepatotoxicity

Name of structural alert	Structural alert	SMARTS pattern
Tamoxifen-like antioestrogen		<chem>[*]N([*])CCOc1ccc(cc1)C(=C\c1ccccc1)\c1ccccc1</chem>
Adenosine-Based Reverse Transcriptase Inhibitors		<chem>[H]N([H])c1ncnc2ncnc12</chem>
Retinoids		<chem>C=CC=CC=CC=O</chem>
b-Lactam substructure		<chem>O=C1CCN1</chem>
Barbitals		<chem>O=C1CC(=O)NC(=O)N1</chem>

Phenothiazines		<chem>[#6]N([#6])CCCN1c2ccccc2Sc2ccccc12</chem>
Cytidine-Based Reverse Transcriptase Inhibitors		<chem>c1cc([N])nc(=O)n1</chem>
Catechols		<chem>[H]Oc1ccccc1O[H]</chem>
Angiotensin-converting enzyme (ACE) inhibitors		<chem>[H]N(CC=O)CC([#7])=O</chem>
Oestrogen Steroids		<chem>[H]Oc1ccc2C3CCC4CCCC4C3CCc2c1</chem>
Thiol		<chem>[H]S[CH3]</chem>
Nitrogen mustards		<chem>ClCCNCCCl</chem>
Anabolic Steroids		<chem>O=C1CCC2C3CCC4CCCC4C3CCC2=C1</chem>
Glucocorticoid steroids		<chem>O=C1C=CC2C3CCC4CCCC4C3CCC2=C1</chem>

Disulfide bridges		[#6]SS[#6]
Para-aminophenyl-sulphonamides		[H]NS(=O)(=O)c1ccc(cc1)N([H])[H]

Procedure Description

Biological Endpoint and Measurement:

This method is an *in silico* model, which provides information on the following, biologically relevant, endpoint:

Hepatotoxicity

Endpoint Value:

Qualitative: Structural alert

Experimental System:

In silico profiler for hepatotoxicity

3. Method Performance

Robustness of the Method

General performance measures:

The performance of each alert was verified by re-screening the original dataset and populating each category. Detailed examination of all category members was undertaken in order to evaluate the performance and to derive a mechanistic hypothesis for the observed hepatotoxicity for each category. Notably, 15 out of 16 categories contained at least one member classified by Fourches et al. (2010) as negative for human hepatotoxicity.

Given the lack of current knowledge into the mechanisms of hepatotoxicity, and the often idiosyncratic nature of the effects, it was accepted from the outset that a mechanistic rationale may not be possible in all cases and some unknown mitigating factors may be responsible for the false positives. Moreover, it is also probable that some hepatotoxic compounds may elicit their toxicities via multiple mechanisms, making the analysis even more difficult (Hewitt et al., 2013).

4. Discussion

The ability of a compound to cause adverse effects to the liver is one of the most common reasons for drug development failures and the withdrawal of drugs from the market. However, the complexity and diversity of hepatotoxicity, the limited (if any) mechanistic insight, the lack of large high quality liver toxicity datasets and the role of metabolism, ensure that hepatotoxicity is one of the most difficult toxicological endpoints to model. The *In silico* profiler for hepatotoxicity of Hewitt et al. (2013) is one of the first attempts undertaken to develop structural alerts or chemical classes for identifying hazardous substances. The study has been followed by other research groups using publicly available hepatotoxicity databases such as Hazard Evaluation Support System (HESS) (Pizzo et al., 2015), Fourches et al. dataset (Fourches et al., 2010; Zhang et al., 2016) or LiverTox human liver injury database (Liu et al., 2015) in development of structural alerts and liver toxicity predictions.

The *in silico* profiler based on mechanistically supported structural alerts provides a useful tool in the screening of compounds during the early stages of their development. Such profiler can be used to identify and group chemicals acting via proposed mechanisms of hepatotoxicity. It also allows for read-across methods to be performed for untested category members if sufficient toxicological data are available.

Alerts of this type are also useful in further exploring the mechanistic basis of hepatotoxicity and have been further corroborated and extended with additional categories developed in subsequent studies of human and rodent hepatotoxicity (eg. Pizzo et al., 2015; Liu et al., 2015; Zhang et al., 2016). A known limitation of the profiler is that some alerts are more specific than others and some may be over-predictive. For example: among 11 barbiturates hepatotoxic in rodents (positive in the structural alert **Barbitals**) only 3 were classified as hepatotoxic in humans (Hewitt et al., 2013). This might be due to species differences or ambiguity in hepatotoxicity definition in the training dataset, as Fourches et al. did not take the exposure into account. Rodent studies typically use higher concentrations than those used in the human clinical applications. Authors caution that the aim of the model is not a stand alone prediction of hepatotoxicity, but to provide a scheme capable of generating mechanistically supported structural alerts suitable for identifying chemicals with hepatotoxicity potential in a weight of evidence approach (Hewitt et al., 2013).

Implemented into an expert system (or weight-of-evidence approach supported by additional data), mechanistically supported structural alerts can be a powerful tool in aiding the identification of hepatotoxic potential. However, taking into account the complexity of hepatotoxicity, the structural alerts are not limited to a single mechanism of action. Indeed, it is probable that many hepatotoxic compounds elicit their toxicities via multiple mechanisms of action. Therefore, it is also possible that the alerts within this profiler may possess the ability to initiate multiple adverse outcome pathways leading to hepatotoxicity and other adverse outcomes.

5. Contact Details

Dr Mark Hewitt

School of Pharmacy and Biomolecular Sciences
Liverpool John Moores University
Method developer / user
Byrom Street
Liverpool L3 3AF

Dr Katarzyna Przybylak

School of Pharmacy and Biomolecular Sciences
Liverpool John Moores University
Method developer / user
Byrom Street
Liverpool L3 3AF

Dr Steve Enoch

School of Pharmacy and Biomolecular Sciences
Liverpool John Moores University

Method developer / user
 Byrom Street
 Liverpool L3 3AF
 email: s.j.enoch@ljmu.ac.uk
 telephone: +44(0)151 231 2164

Prof. Mark Cronin
 Professor of Predictive Toxicology
 School of Pharmacy and Biomolecular Sciences
 Liverpool John Moores University
 Corresponding author
 Byrom Street
 Liverpool L3 3AF
 email: m.t.cronin@ljmu.ac.uk
 telephone: + 44 151 231 2402
 fax: + 44 151 231 2170
 last contact update: 11 May 2016

Ver n°	Date	Name	Type of change
1	12 November 2015	Prof. Mark Cronin	Data sheet created

6. Abbreviations & Definitions

AOP Adverse Outcome Pathway

COSMOS EU Integrated Project COSMOS developed an integrated suite of computational tools to predict the effects of long-term exposure to chemicals in humans based on *in silico* calculation and estimation techniques.

Website: <http://www.cosmostox.eu>

DILI Drug-induced liver injury

ITS Integrated Testing Strategy

IATA Integrated Approach to Testing and Assessment

KNIME KoNstanz Information MinEr

MIE Molecular Initiating Event

OECD Organisation for Economic Cooperation and Development

(Q)SAR (Quantitative) Structure-Activity Relationships

SEURAT Safety Evaluation Ultimately Replacing Animal Testing

SMARTS SMiles ARbitrary Target Specification

SMILES Simplified Molecular-Input Line-Entry System

7. Bibliography

Method development:

- Enoch S.J., Cronin M.T.D., Ellison C.M. (2011)
 The use of a chemistry based profiler for covalent DNA binding in the development of chemical categories for read-across for genotoxicity.
Alternatives to Laboratory Animals (ATLA) 39, 131-145
- Fourches D., Barnes J.C., Day N.C., Bradley P., Reed J.Z., Tropsha A. (2010)
 Cheminformatics analysis of assertions mined from literature that describe drug-induced liver injury in different species
Chemical Research in Toxicology 23, 171-183

- Hewitt M., Enoch S.J., Madden J.C., Przybylak K.P., and Cronin M.T.D. (2013)
Hepatotoxicity: A scheme for generating chemical categories for read-across, structural alerts and insights into mechanism(s) of action.
Critical Reviews in Toxicology 43(7), 537-558

General references:

- Landesmann B. (2016)
AOP 38: Protein Alkylation leading to Liver Fibrosis. Access: <https://aopwiki.org/wiki/index.php/Aop:38>
- Landesmann B. (2012)
Description of Prototype Modes-of-Action Related to Repeated Dose Toxicity,
JRC publication repository EUR 25631 EN

Last update: 12 November 2015

DB-ALM Method summary n° 180 : *In silico* profiler for mitochondrial toxicity

Systemic Toxicity

The *in silico* profiler for mitochondrial toxicity consists of 21 structural alerts coded into SMARTS patterns. This profiler is envisaged to be used to develop chemical categories based upon similar mechanisms as part of the adverse outcome pathway paradigm (AOP). Additionally, the profiler could be utilised in screening large data sets in order to identify chemicals with the potential to induce mitochondrial toxicity.

1. General Information

Corresponding author(s):

Dr Steve Enoch, Liverpool John Moores University

Date of data sheet creation: 12 November 2015

Last version number: 1

Last update: 12 November 2015

Proprietary and/or Confidentiality Issues:

None. The model is publicly available.

Status

Development

Work carried out at Liverpool John Moores University with support from the FP7 COSMOS Project (Integrated *In Silico* Models for the Prediction of Human Repeated Dose Toxicity of Cosmetics to Optimise Safety), within the SEURAT-1 Research Initiative (Safety Evaluation Ultimately Replacing Animal Testing 1) cluster (2011-2015).



The *in silico* profiler for mitochondrial toxicity has a number of structural alerts coded into SMARTS patterns and translated into KNIME (<http://www.knime.org/>) automated workflows as part of the COSMOS project. The KNIME workflow is available via COSMOS Space

<http://cosmospace.cosmostox.eu>

A web version of the model was also implemented, allowing to execute the model via a web browser. It is available through the COSMOS KNIME WebPortal:

<http://knimewebportal.cosmostox.eu>

2. Method Definition

Purpose of the Method

TYPE OF TESTING	:	Screening, Part of an integrated testing strategy
LEVEL OF TOXICITY ASSESSMENT	:	Toxic potential

Context of use

The ability to predict organ-level toxicity is becoming increasingly important to the long term goal of replacing animal use or chemical safety assessment. Many organ-specific and systemic toxicity pathways can be traced back to mitochondrial function as initial target (Dykens and Will, 2007). Therefore, there is an urgent need to develop models able to early detect potential mitochondria toxicants.

The main goal of the *in silico* profiler is grouping chemicals into mechanism based categories centred on the knowledge of a number of key molecular initiating events (MIEs) for mitochondrial toxicity as part of

the adverse outcome pathway (AOP) paradigm (e.g. Event 177: "*Mitochondrial dysfunction*"; AOP-Wiki, 2016). Such categories allow for read-across and data gap filling to be applied. The premise behind these structurally developed categories is that similar chemicals should have similar biological activities and therefore should have the same MIE.

Furthermore, the categories produced using *in silico* profilers can be supported by, and - screening large data sets in order to identify chemicals with potential mitochondrial toxicity - used to prioritise, additional testing using *in vitro* and/or *in chemico* methods, within an integrated testing strategy (ITS) or an integrated approach to testing and assessment (IATA) (Tollefsen et al., 2014). Such strategies can be used in hazard identification and risk assessment for regulatory purposes such as REACH (EU, 2006) or Cosmetics Regulation (EU, 2009) and could be incorporated into *in silico* decision-supporting software tools such as the OECD QSAR Toolbox.

Scientific Principle of the Method

One of the most ubiquitous and yet less studied targets of systemic toxicity is mitochondrial dysfunction, which has been linked to a variety of organ toxicities within kidney, liver and nervous tissues (Dyken and Will 2007).

The molecular mechanism of mitochondrial toxicity is complex: disruption of mitochondrial DNA replication, protein synthesis, fatty acid metabolism or Krebs cycle, inhibition of electron transport, uncoupling of electron transport from ATP synthesis, membrane permeability transition and production of reactive oxygen species are among possible mechanisms which can be related to mitochondrial dysfunction (Dyken and Will 2007). In addition, mode of action studies reveal that xenobiotics may cause mitochondrial disruption through several mechanisms simultaneously (Nelms et al., 2015).

Within the AOP paradigm, *in silico* methods contribute to identifying the key features of a chemical (structural alerts) that induce a MIE. A collection of structural alerts that induce the same MIE are considered to be an *in silico* profiler. There are two types of profilers: mechanism-based profilers and chemistry-based profilers. The first type is comprised of structural alerts relating to MIE that are supported by experimental data illustrating how the alerts initiate the MIE, for example the *in silico* profilers for covalent protein binding (DB-ALM Method Summary No. 181), covalent DNA binding (DB-ALM Method Summary No. 178) or mitochondrial toxicity described in this Method Summary.

In comparison, the chemistry-based profiler consists of structural alerts that have been associated with inducing toxicity; thus, in contrast to mechanism-based profilers, a chemistry-based profiler does not contain mechanistic information relating to how the observed toxicity is initiated. Examples are: the *in silico* profilers for hepatotoxicity (DB-ALM Method Summary No. 179) and for potential nuclear receptor ligands and alerts towards hepatosteatosis (DB-ALM Method Summary No. 177).

Due to these intrinsic differences, the two types of profiler should be used for different purposes; mechanistic profilers should be used for category formation and read-across, whilst chemistry-based profilers should be used to screen an inventory to prioritise chemicals to undergo (non-animal) testing. A category developed utilising a mechanism-based profiler enables missing toxicological data for a chemical to be filled using available information from analogues within the same category via read-across predictions (Enoch et al., 2011).

Model Description

The *in silico* profiler for mitochondrial toxicity was developed based around clearly defined mechanistic information. This was achieved by grouping chemicals based upon their structural similarity, followed by a literature search to elucidate mechanistic information for a number of key molecular initiating events that disrupt the normal functioning of mitochondria.

The identified structural alerts were defined as SMARTS (SMiles ARbitrary Target Specification, based on SMILES - Simplified Molecular-Input Line-Entry System) patterns. SMARTS pattern allows for a definition of a sub-structure within a molecule, which is then used to check whether a chemical structure contains any of the pre-defined fragments.

The *in silico* profiler was implemented as KNIME workflow, and additionally as web version, allowing to execute the model via a web browser. It is available through the COSMOS KNIME WebPortal:

<http://knimewebportal.cosmostox.eu>

Datasets

Two datasets were used for the development of structural alerts associated with mitochondrial toxicity:

1. 288 chemicals reported by Zhang et al (2009). This article was chosen as it provides one of the largest freely available datasets, for which the chemical list has qualitative mitochondrial toxicity data associated with it. Within this data set 171 chemicals have been reported within the literature as inducing mitochondrial toxicity and were therefore considered to be mitochondrial toxicants. The chemicals with a negative result for mitochondrial toxicity were selected from the FDA-approved drug list, whereby the therapeutic action mechanism, of common and safe oral drugs, was not associated with a mechanism of drug-induced mitochondrial toxicity. Data for the 288 chemicals, including SMILES and toxicity towards mitochondria are available in Nelms et al.(2015 bis).
2. 94 hair dyes with NOAEL values from oral 90-day rat studies extracted from the SCC(NF)/SCCS 'Opinion On' reports published between 2000 and 2013. The dataset including SMILES and NOAEL values is available in Nelms et al. (2015).

Category formation based upon structural similarity

All chemical structures were encoded into SMILES strings and neutralised, and salts were removed. Similarity calculations were implemented within the freely available Toxmatch software (v1.07) using the atom environment nearest-neighbour approach, generating a data matrix with a Tanimoto similarity score for each chemical to all others within the dataset. Subsequently, in-house code was implemented within Microsoft Excel that identified analogues with a similarity index of 0.6 or greater; this was used to develop categories for the chemicals within the dataset with two, or more, analogues. Further analysis was undertaken upon those categories that met the following criteria: they contained three or more chemicals and at least one mitochondrial toxic chemical.

In total, twelve categories were initially formed:

1. local anaesthetics, antianginal and antiarrhythmic;
2. anti-diabetic drugs;
3. nonsteroidal anti-inflammatory drugs;
4. anthracycline antibiotics;
5. hypolipodemic drugs;
6. bile acids;
7. antihistaminic, antipsychotic and antiemetic drugs;
8. β -blockers;
9. 2-nitroaminobenzenes;
10. 4-nitroaminobenzenes;
11. aromatic azos and
12. anthraquinones.

Mechanistic hypothesis and development of alerts

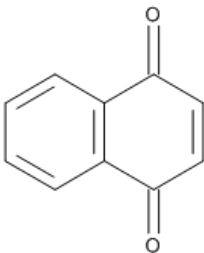
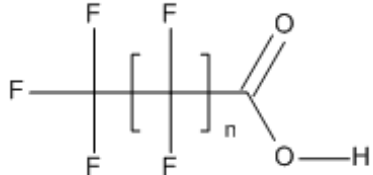
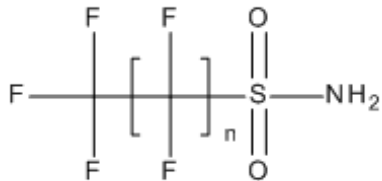
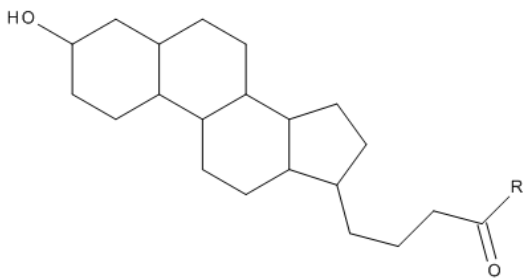
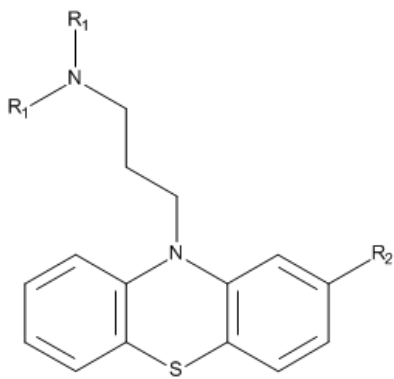
Once categories had been developed using structural similarity, a detailed search of the available literature was undertaken to elucidate the mechanistic information behind the molecular initiating event, along with other downstream key events, leading to the disruption of the mitochondria. This mechanistic information was subsequently utilised to support the definition of a structural alert suitable for grouping chemicals. These structural alerts were defined by identifying the common fragment present within each of the chemicals found to have positive mitochondrial toxicity according to literature information associated with them. Any additional information regarding the limits of the fragment found during the literature search, such as the requirement for an electron-withdrawing group or the type of bond needed (e.g., a tertiary amine), was used to further refine the structural alert.

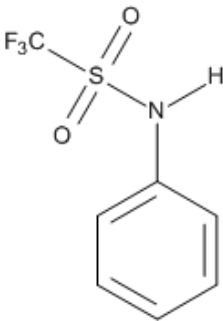
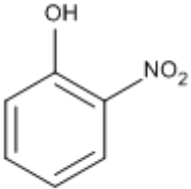
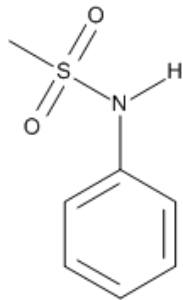
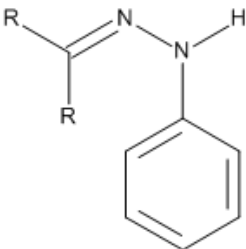
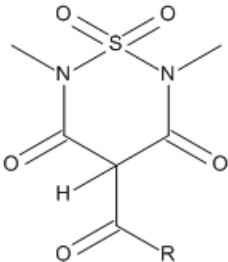
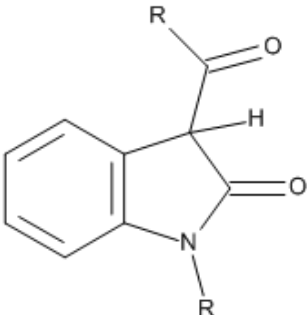
As a result, 21 structural alerts were developed (nine of them were earlier identified by Naven et al. (2013)) and subsequently defined as SMARTS patterns.

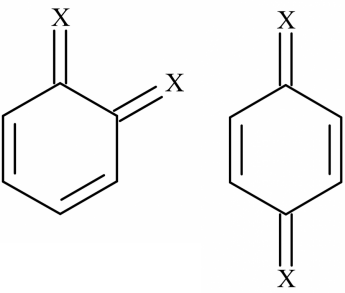
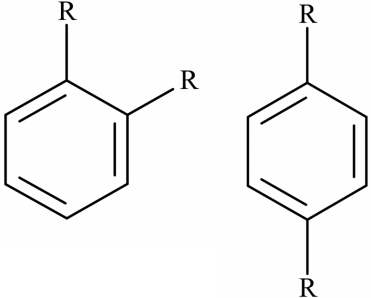
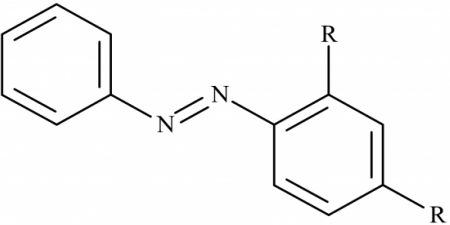
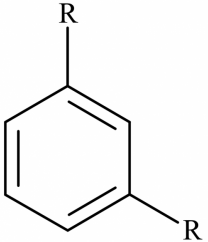
Table 1 presents the list of 21 structural alerts together with the underlying mechanism and their graphical representation. The SMARTS pattern are available in **Table 2**. The full description of this profiler and the mechanistic rationale behind each alert is available in Nelms et al. (2015 and 2015 bis).

Table 1. Summary of the structural alerts identified for mitochondrial toxicity.

Structural alert	Mechanism	Structural Alert
2-(Dialkylamino)-N-(2,6-dimethylphenyl)acetamides	Protonophore	<p>R = alkyl carbon</p>
Thiazolinediones	Protonophore	
2-Anilinobenzoic acids	Protonophore	
Salicylates	Protonophore	
Anthranilic acids	Protonophore	

Anthracene-9,10-diones	Redox cycling	
Naphthoquinones	Redox cycling	
Perfluorinated-carboxylic-acids	Protonophore	 <p>n = 5-11</p>
Perfluorinated-sulphonamides	Protonophore	 <p>n = 5-11</p>
Bile acids	Unclear	 <p>R = alkyl carbon</p>
Phenothiazines	Inhibition of Complexes I-IV	 <p>R₁ = CH₂, CH₃ R₂ = Cl, CF₃</p>

Fluoromethylsulphanilides	Protonophore	
Nitrophenols	Protonophore	
Nitrosulphonanilides	Protonophore	
Phenylhydrazones	Protonophore	 <p>R = any carbon</p>
Thiadiazinedione dioxides	Protonophore	
Acyindolones	Protonophore	 <p>R = alkyl carbon</p>

Quinones (X = NH, O)	Inhibition of the electron transport chain and uncoupling of oxidative phosphorylation	
Pro-quinones (R = OH, NH ₂ , NH, NO ₂)	Inhibition of the electron transport chain and uncoupling of oxidative phosphorylation	
Aromatic azos (R = NH ₂ , NH, OH)	Inhibition of the electron transport chain	
Meta-substituted benzenes (R = NH ₂ , OH)	Inhibition of the electron transport chain	

Aromatic azos (R = NH ₂ , NH, OH)	<chem>\$(a[NX2]=[NX2]c1c([NH2,NH,OH])cccc1),\$(a[NX2]=[NX2]c1ccc([NH2,NH,OH])cc1)</chem>
Meta-substituted benzenes (R = NH ₂ , OH)	<chem>c1ccc([NH2,NH,OH])c(!\$([NH2,NH,OH]))c1[NH2,NH,OH]</chem>

Procedure Description

Biological Endpoint and Measurement:

This method is an *in silico* model, which provides information on the following, biologically relevant, endpoint:

MITOCHONDRIAL TOXICITY: Structure-activity relationships related to protonophores, redox cycling, complex I-IV inhibition, inhibition of electron transport chain and uncoupling of oxidative phosphorylation.

Endpoint Value:

Qualitative: Structural alert

Experimental System:

In silico profiler for structural alerts of mitochondrial toxicity

3. Method Performance

Robustness of the Method

General performance measures:

The method detects structure-activity relationships with a well documented mechanistic basis for mitochondrial toxicity, however the profiler is not exhaustive and no analysis of each alert's performance has been conducted yet.

An earlier study by Naven et al. (2013) has shown that chemicals which share the same structural alert, may have different mitochondrial dysfunction potency. Therefore the presence of an alert should be considered an indication of potential mitochondrial toxicity that can be used for screening and a guidance for further *in vitro* studies for mitochondrial dysfunction.

4. Discussion

Strengths and limitations of the method

The ability to predict organ-level toxicity is becoming increasingly important to the long term goal of replacing animal use for chemical safety assessment. It is generally understood that this goal can only be achieved by a combination of complementary approaches (e.g. *in vitro*, *ex vivo*, *in silico*, *in chemico*) that address functional mechanistic endpoints tied to adverse outcomes of regulatory concern.

Organ toxicities within kidney, liver and nervous tissues have been frequently linked to mitochondrial dysfunction, causing for example off-target systemic adverse effects in chronic drug exposure (Dyken and Will, 2007). Such effects are often missed in the pre-clinical drug safety animal studies and may be driven by metabolic or genetic components present only in specific patient subpopulations. In their

review in 2007 Dykens and Will estimated that approximately 50% of drugs with FDA Black Box warning for hepatotoxicity and cardiovascular toxicity cause mitochondrial dysfunction. Consequently mitochondrial dysfunction is regarded as MIE in numerous AOPs. This example from drug safety assessment is also applicable for general chemical safety assessment and shows that there is an urgent need to develop better models able to early detect potential human mitochondria toxicants.

The structural alerts demonstrate such ability for identifying chemicals which can disrupt the mitochondrial functionality. The *in silico* profiler can be utilised to either screen an inventory for prioritization or to develop chemical categories, from which read-across predictions could be made regarding to a chemical's ability to initiate mitochondrial toxicity and for which further *in vitro* / *in chemico* testing may be appropriate.

Upon development of structural alerts it is necessary to substantiate that each of the alerts correctly identifies chemicals with the potential to instigate an MIE. Other alternative techniques, such as *in vitro* and/or *in chemico* assays, can be used to verify the alerts developed are correct. However, it should be noted that whilst these alerts may be correct, other factors, such as the internal concentration or metabolism, may prevent the MIE from being induced *in vivo*.

It is important to mention that a structural alert was developed only if information linking category members to mitochondrial toxicity was present within the scientific literature. The benefit of undertaking the analysis for each category is that it enabled the chemical space associated with a known, and tested, mechanism of mitochondrial toxicity to be identified. The development of chemical categories and identification of additional mechanistic information from the literature were crucial in addressing the limitations of the information in the original data set. It should be noted that, due to the limitations of the data available, the method does not aim to create a profiler that would be used for pharmaceutical risk assessment. Its development has highlighted the need for more mitochondrial chemical testing and better understanding of the mitochondrial dysfunction and its role in adverse health effects (Nelms et al., 2015 bis).

Potential for future development

Additional future work involves utilising experimental information from *in vitro* / *in chemico* assays to verify the mechanism-based alerts and to refine the chemistry-based alerts by discerning mechanistic information associated with them.

5. Contact Details

Prof. Mark Cronin

Professor of Predictive Toxicology
School of Pharmacy and Biomolecular Sciences
Liverpool John Moores University
Method developer / user
Byrom Street
Liverpool L3 3AF
email: m.t.cronin@ljmu.ac.uk
telephone: + 44 151 231 2402
fax: + 44 151 231 2170
last contact update: 11 May 2016

Dr Mark Nelms

School of Pharmacy and Biomolecular Sciences
Liverpool John Moores University
Method developer / user
Byrom Street
Liverpool L3 3AF

Dr Claire Mellor

School of Pharmacy and Biomolecular Sciences
Liverpool John Moores University
Method developer / user
Byrom Street
Liverpool L3 3AF

Dr Judith Madden

School of Pharmacy and Biomolecular Sciences
 Liverpool John Moores University
 Method developer / user
 Byrom Street
 Liverpool L3 3AF

Dr Steve Enoch

School of Pharmacy and Biomolecular Sciences
 Liverpool John Moores University
 Corresponding author, Method developer / user
 Byrom Street
 Liverpool L3 3AF
 email: s.j.enoch@ljmu.ac.uk
 telephone: +44(0)151 231 2164

Prof Vera Rogiers

Department of Toxicology, Center for Pharmaceutical Research (CePhar)
 Vrije Universiteit Brussel
 Method developer / user

Dr Gamze Ates

Department of Toxicology, Center for Pharmaceutical Research (CePhar)
 Vrije Universiteit Brussel
 Method developer / user

Dr Mathieu Vinken

Department of Toxicology, Center for Pharmaceutical Research (CePhar)
 Vrije Universiteit Brussel
 Method developer / user

Ver n°	Date	Name	Type of change
1	12 November 2015	Dr Steve Enoch	Datasheet created

6. Abbreviations & Definitions

AOP Adverse Outcome Pathway

COSMOS EU Integrated Project: COSMOS developed an integrated suite of computational tools to predict the effects of long-term exposure to chemicals in humans based on *in silico* calculation and estimation techniques.

Project website: <http://www.cosmostox.eu>.

ITS Integrated Testing Strategy

IATA Integrated Approach to Testing and Assessment

KNIME KoNstanz Information MinEr

LO(A)EL Lowest Observed (Adverse) Effect Level

MIE Molecular Initiating Event

OECD Organisation for Economic Cooperation and Development

(Q)SAR (Quantitative) Structure-Activity Relationships

SCCNFP Scientific Committee on Cosmetic products and Non-Food Products

SCCP Scientific Committee on Consumer Products

SCCS Scientific Committee on Consumer Safety

SEURAT Safety Evaluation Ultimately Replacing Animal Testing

SMARTS SMiles ARbitrary Target Specification

SMILES Simplified Molecular-Input Line-Entry System

7. Bibliography

Method development:

- Nelms M.D., Ates G., Madden J.C., Vinken M., Cronin M.T.D, Rogiers V., Enoch S.J. (2015)
Proposal of an in silico profiler for categorisation of repeat dose toxicity data of hair dyes.
Archives of Toxicology 89, 733-741
- Nelms M.D., Mellor C.L., Cronin M.T.D, Madden J.C., and Enoch S.J. (2015)
The development of an in silico profiler for mitochondrial toxicity.
Chemical Research in Toxicology 28, 1891-1902
- Zhang H., Chen Q.Y., Xiang M.L., Ma C.Y., Huang Q., Yang S.Y (2009)
In silico prediction of mitochondrial toxicity by using GA-CG-SVM approach.
Toxicology In Vitro 23, 134-140

General references:

- AOP-Wiki (2016)
Event 177: Mitochondrial dysfunction (<https://aopwiki.org/wiki/index.php/Event:177> Last access 13.09.2016)
Collaborative Adverse Outcome Pathway Wiki (AOP-Wiki)
- Dykens J.A. and Will Y. (2007)
The significance of mitochondrial toxicity testing in drug development.
Drug Discovery Today 12, 77-785
- EU (2009)
Regulation (EC) No 1223/2009 of the European Parliament and of the Council of 30 November 2009 on cosmetic products. [Link to document](#) (last access 13.09.2016)
Official Journal of the European Communities
- EU (2006)
Regulation (EC) No 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH). [Link to document](#) (last access 16.09.2013)
Official Journal of the European Communities L396
- Enoch S.J., Cronin M.T.D., Ellison C.M. (2011)
The use of a chemistry based profiler for covalent DNA binding in the development of chemical categories for read-across for genotoxicity.
Alternatives to Laboratory Animals (ATLA) 39, 131-145
- Naven R.T., Swiss R., Klug-Mcleod J., Will Y., Greene N. (2013)
The development of structure-activity relationships for mitochondrial dysfunction: Uncoupling of oxidative phosphorylation.
Toxicological Sciences 131, 271-278
- Tollefsen K.E., Scholz S., Cronin M.T., Edwards S.W., de Knecht J., Crofton K., Garcia-Reyero N., Hartung T., Worth, A., Patlewicz G. (2014)
Applying Adverse Outcome Pathways (AOPs) to support Integrated Approaches to Testing and Assessment (IATA).
Regulatory Toxicology and Pharmacology 70(3), 629-640

Last update: 12 November 2015

DB-ALM Method summary n° 181 : *In silico* profiler for covalent protein binding

Systemic Toxicity

The *in silico* profiler for covalent protein binding compiles mechanistic organic chemistry fragments (in the form of structural alerts) for the binding of exogenous chemicals (electrophiles) to proteins (nucleophiles). The profiler is envisaged to be used to develop chemical categories based upon similar mechanisms as part of the adverse outcome pathway paradigm. Additionally, the profiler could be utilised in screening large data sets in order to identify chemicals with the potential to bind to proteins.

1. General Information

Corresponding author(s):

Dr Steve Enoch, Liverpool John Moores University

Date of data sheet creation: 12 November 2015

Last version number: 1

Last update: 12 November 2015

Proprietary and/or Confidentiality Issues:

None. The model is publicly available.

Status

Development

Development completed. Work carried out at Liverpool John Moores University with support from the with support from the European Chemicals Agency (ECHA) Service Contract No. ECHA/2008/20/ECA.203 and the FP7 COSMOS Project (Integrated *In Silico* Models for the Prediction of Human Repeated Dose Toxicity of Cosmetics to Optimise Safety), within the SEURAT-1 (Safety Evaluation Ultimately Replacing Animal Testing 1) cluster (2011-2015). More information on the COSMOS project is available on its dedicated website:

<http://www.cosmostox.eu>



Most of the structural alerts and associated mechanistic chemistry have been incorporated into the Organisation for Economic Co-operation and Development (OECD) (Q)SAR Application Toolbox, a software tool designed to fill data gaps in a regulatory environment without the need for further animal testing: <http://www.qsartoolbox.org>

The *in silico* profiler for covalent protein binding has a number of structural alerts coded into SMARTS patterns and translated into KNIME (<http://www.knime.org/>) automated workflows as part of the COSMOS project. The KNIME workflow is available via COSMOS Space:

<http://cosmospace.cosmostox.eu>

A web version of the model was also implemented, allowing to execute the model via a web browser. It is available through the COSMOS KNIME WebPortal:

<http://knimewebportal.cosmostox.eu>

2. Method Definition

Purpose of the Method

TYPE OF TESTING	: Screening, Part of an integrated testing strategy
LEVEL OF TOXICITY ASSESSMENT	: Toxic potential

Context of use

The main goal of the *in silico* profiler is grouping chemicals into mechanism based categories centred on the knowledge of a number of key molecular initiating events (MIEs) for a number of toxicity endpoints such as skin sensitisation, respiratory sensitisation, hepatotoxicity and aquatic toxicity, as part of the adverse outcome pathway (AOP) paradigm.

Such categories allow for read-across and data gap filling to be applied to the chemicals for which no toxicological data are available. The premise behind these structurally developed categories is that similar chemicals should have similar biological activities and therefore should have the same MIE.

Covalent binding to extra- and intra-cellular proteins is a well known mechanism of action of chemicals, leading to impaired physiological functions of the modified proteins and/or immune response to the haptenated proteins (Zhou et al., 2005; Chipinda et al., 2011). Target organ or systemic adverse effects may include allergic contact dermatitis, respiratory sensitisation, liver fibrosis, neurotoxicity, cancer, growth impairment or mortality of aquatic organisms (e.g. AOP 40: "Covalent Protein binding leading to Skin Sensitisation" and AOP 38: "Protein Alkylation leading to Liver Fibrosis") (Landesmann, 2012 and 2016; Lopachin and Decaprio, 2005). Such effects are of major concern in drug development and for the environmental, health and safety impact of chemicals.

The categories produced using *in silico* profilers can be supported by -and screening large data sets in order to identify chemicals with the potential to bind to proteins, and can be used to prioritise additional testing using *in vitro* and/or *in chemico* methods, within an integrated testing strategy (ITS) or an integrated approach to testing and assessment (IATA).

Such strategies can be used in hazard identification and risk assessment for regulatory purposes such as REACH (EU, 2006) or the Cosmetics Regulation (EU, 2009). To facilitate the use in this context the covalent DNA binding profiler was incorporated into *in silico* software tools such as the OECD QSAR Toolbox.

Scientific Principle of the Method

For a number of toxicological endpoints, such as hepatotoxicity (liver fibrosis), skin sensitisation, respiratory sensitisation or aquatic toxicity, one of the fundamental steps is the formation of a covalent adduct between a reactive exogenous chemical (electrophile) and biological protein (nucleophile). Protein binding can thus be a molecular initiating event for a number of adverse outcome pathways (e.g. AOP 40 and AOP 38, Landesmann, 2016; Cochrane et al., 2015).

Within the AOP paradigm, *in silico* methods contribute to identifying the key features of a chemical (structural alerts) that induce a MIE. A collection of structural alerts that induce the same MIE are considered to be an *in silico* profiler.

There are two types of profilers: mechanism-based profilers as described here and chemistry-based profilers. The first type is comprised of structural alerts relating to a MIE that are supported by experimental data illustrating how the alerts initiate the MIE, as for example also the *in silico* profilers for covalent DNA binding (DB-ALM Method Summary No. 178) or mitochondrial toxicity (DB-ALM Method Summary No. 180). In comparison, the chemistry-based profiler, as for example the *in silico* profilers for potential nuclear receptor ligands and alerts towards hepatosteatosis (DB-ALM Method Summary No. 177) or hepatotoxicity (DB-ALM Method Summary No. 179), consists of structural alerts that have been associated with inducing toxicity; thus, in contrast to mechanism-based profilers, a chemistry-based profiler does not contain mechanistic information relating to how the observed toxicity is initiated.

Due to these intrinsic differences, the two types of profilers should be used for different purposes; mechanistic profilers should be used for category formation and read-across, whilst chemistry-based profilers should be used to screen an inventory to prioritise chemicals to undergo (non-animal) testing. A category developed utilising a mechanism-based profiler enables missing toxicological data for a chemical to be filled using available information from analogues within the same category via read-across predictions (Enoch et al., 2011).

Model Description

The *in silico* profiler for covalent protein binding was developed based on the review of current scientific knowledge on structural alerts relating to a number of toxicity endpoints such as skin sensitisation, respiratory sensitisation and aquatic toxicity. There are several publications in which structural alerts for direct acting and indirect acting electrophiles have been published (see Enoch et al., 2011 bis).

The existing structural alerts were mapped in terms of their relationships with mechanistic organic chemistry, i.e. identifying alerts from the published compilations related to covalent protein binding. The mapping was performed to achieve maximum overlap and usability whilst restricting redundancy in the alerts, and to ensure that the alerts related to the molecular initiating event of covalent protein binding.

As a result, a total of 108 structural alerts, supported by mechanistic information and meta data, have been created, covering five broad organic chemistry mechanisms (domains) - acylation (AC), Michael addition (MA), Schiff base formation (SB), bimolecular aliphatic nucleophilic substitution (S_N2) and aromatic nucleophilic substitution (S_NAr).

The identified mechanistic structural alerts were further defined as SMARTS (SMiles ARbitrary Target Specification, based on SMILES - Simplified Molecular-Input Line-Entry System) patterns. SMARTS pattern allows for a definition of a sub-structure within a molecule, which is then used to check whether a chemical structure contains any of the pre-defined fragments.

Moreover, the detailed mechanistic chemistry reaction associated with each of the alerts has been compiled. The mechanistic information is intended to outline how the alert can act as a direct electrophile or how it can be converted into an electrophile. Therefore, an important consideration within the mechanistic chemistry framework is the inclusion of potential metabolic activation.

The *in silico* profiler was implemented as KNIME workflow, and additionally as web version, allowing to execute the model via a web browser. It is available through the COSMOS KNIME WebPortal:

<http://knimewebportal.cosmostox.eu>

Table 1. Summary of the 108 structural alerts defined for covalent protein binding.

AC - Acylation, MA - Michael addition, SB - Schiff base formation,
 S_N2 - nucleophilic substitution 2, S_NAr - aromatic nucleophilic substitution

The full description of the protein binding profiler is available in Enoch et al. (2011 bis).

Structural alert	SMARTS	Mechanism
Acyl halides AC	<chem>[CH,\$(C(=[O,SX1])[#6]))(=[O,SX1])[F,Cl,Br,I]</chem>	Direct acylation involving a leaving group
Acetates AC	<chem>[CH,\$(C(=[O,SX1])[#6]))(=[O,SX1])[O,SX2,NX3] \$([a;r6]),\$([a;r5]),\$(C([O,SX2,NX3])=C), \$(C([O,SX2,NX3])#C)]</chem>	Direct acylation involving a leaving group
Anhydrides AC	<chem>[CH,\$(C([O,SX2])(=[O,SX1])[#6]))(=[O,SX1])[O,SX2] [CH,\$(C([O,SX2])(=[O,SX1])[#6]))=[O,SX1]</chem>	Direct acylation involving a leaving group
Azlactones AC	<chem>C1=[O,SX2,NX2,CX3]CC(=[O,SX2])[O,SX2,NX3]1</chem>	Direct acylation involving a leaving group
Sulphonyl halides AC	<chem>[#6]S(=O)(=O)[F,Cl,Br,I,\$(C#N)]</chem>	Direct acylation involving a leaving group
Phosphonic acid halides AC	<chem>[#6]OP(=O)(O[#6])[F,Cl,Br,I,\$(C#N)]</chem>	Direct acylation involving a leaving group
Dialkyl carbamyl halides AC	<chem>[NH2,\$([NH](C=O)[#6]),\$(N(C=O)([#6])[#6]))C(=O)[F,Cl,Br,I]</chem>	Direct acylation involving a leaving group
β -Lactones AC	<chem>C1C(=O)OC1</chem>	Ring opening acylation
Thio lactones AC	<chem>C1C(=O)SC1</chem>	Ring opening acylation
α -Lactams AC	<chem>C1C(=O)NC1</chem>	Ring opening acylation

Cyclopropenones AC	<chem>[cH,\$(c(c)[#6])]1[cH,\$(c(c)[#6])]c1=O</chem>	Ring opening acylation
Isocyanates AC	<chem>[NH,\$(N(=C=O)[#6])]=C=O</chem>	Isocyanates and related
Isothiocyanates AC	<chem>[NH,\$(N(=C=[SX1])[#6])]=C=[SX1]</chem>	Isocyanates and related
Dithiocarbonamide acid esters AC	<chem>[#6]N=C(S[#6])S[#6]</chem>	Isocyanates and related
Carbodiimides AC	<chem>[#6]N=C=N[#6]</chem>	Isocyanates and related
Ketenes AC	<chem>[CH2,\$([CH](=C=O)[#6]),\$(C(=C=O)([#6])[#6])]=C=O</chem>	Isocyanates and related
Polarised alkene aldehydes MA	<chem>[\$([CH2]),\$([CH][#6])]=[CH,\$(C[#6])][CH]=O</chem>	Polarised alkenes
Polarised alkene ketones MA	<chem>[\$([CH2]),\$([CH][#6])]=[CH,\$(C[#6])C(=O)[#6]);!\$([CH]1=CC(=O)C=CC1(=O));!\$([CH]1=CC(=O)C(=O)C=C1)]</chem>	Polarised alkenes
Polarised alkene esters MA	<chem>[\$([CH2]),\$([CH][#6])]=[CH,\$(C[#6])C(=O)O[#6]</chem>	Polarised alkenes
Polarised alkene amides MA	<chem>[\$([CH2]),\$([CH][#6])]=[CH,\$(C[#6])C(=O)NH2,\$([NH](C=O)[#6]),\$(N(C=O)([#6])[#6])]</chem>	Polarised alkenes
Polarised alkene nitros MA	<chem>[\$([CH2]),\$([CH][#6])]=[CH,\$(C[#6])][N+](=O)[O-]</chem>	Polarised alkenes
Polarised alkene cyano MA	<chem>[\$([CH2]),\$([CH][#6])]=[CH,\$(C[#6])C#N</chem>	Polarised alkenes
Polarised alkene sulfonate MA	<chem>[\$([CH2]),\$([CH][#6])]=[CH,\$(C[#6])S(=O)(=O)O[#6]</chem>	Polarised alkenes
Polarised alkene sulfone MA	<chem>[\$([CH2]),\$([CH][#6])]=[CH,\$(C[#6])S(=O)(=O)[#6]</chem>	Polarised alkenes
Polarised alkene sulfinyl MA	<chem>[\$([CH2]),\$([CH][#6])]=[CH,\$(C[#6])][SX3](=O)[#6]</chem>	Polarised alkenes
Polarised alkene oximes MA	<chem>[\$([CH2]),\$([CH][#6])]=[CH,\$(C[#6])C=[NX2][OH]</chem>	Polarised alkenes
Polarised alkene pyridines MA	<chem>[\$([CH2]),\$([CH][#6])]=[CH,\$(C[#6])][\$(c1ncccc1),\$(c1ccncc1)]</chem>	Polarised alkenes
Polarised alkene pyrazines MA	<chem>[\$([CH2]),\$([CH][#6])]=[CH,\$(C[#6])c1nccnc1</chem>	Polarised alkenes
Polarised alkene pyrimidines MA	<chem>[\$([CH2]),\$([CH][#6])]=[CH,\$(C[#6])][\$(c1ncncc1),\$(c1ncccn1),\$(c1cncnc1)]</chem>	Polarised alkenes
Polarised alkene triazines MA	<chem>[\$([CH2]),\$([CH][#6])]=[CH,\$(C[#6])][\$(c1ncncc1),\$(c1nnccn1),\$(c1nnccn1),\$(c1cnncn1),\$(c1nnncc1),\$(c1cnncn1)]</chem>	Polarised alkenes
Azocarbonamides MA	<chem>[NH2,\$([NH](C=O)[#6]),\$(N(C=O)([#6])[#6])C(=O)N=NC(=O)NH2,\$([NH](C=O)[#6]),\$(N(C=O)([#6])[#6])]</chem>	Polarised azo compounds
Polarised alkyne aldehydes MA	<chem>[CH,\$(C(#C)[#6])][#C][CH]=O</chem>	Polarised alkynes
Polarised alkyne ketones MA	<chem>[CH,\$(C(#C)[#6])][#CC(=O)[#6]</chem>	Polarised alkynes
Polarised alkyne esters MA	<chem>[CH,\$(C(#C)[#6])][#CC(=O)O[#6]</chem>	Polarised alkynes

Polarised alkyne amides MA	[CH,\$(C(#C)[#6])#CC(=O)[\$(NH2)], \$([NH](C=O)[#6]),\$(N(C=O)([#6])[#6])]	Polarised alkynes
Polarised alkyne nitros MA	[CH,\$(C(#C)[#6])#C[N+](=O)[O-]	Polarised alkynes
Polarised alkyne cyano MA	[CH,\$(C(#C)[#6])#CC#N	Polarised alkynes
Polarised alkyne sulfonate MA	[CH,\$(C(#C)[#6])#CS(=O)(=O)O[#6]	Polarised alkynes
Polarised alkyne sulfone MA	[CH,\$(C(#C)[#6])#CS(=O)(=O)[#6]	Polarised alkynes
Polarised alkyne sulfinyl MA	[CH,\$(C(#C)[#6])#C[SX3](=O)[#6]	Polarised alkynes
Polarised alkyne pyridines MA	[CH,\$(C(#C)[#6])#C\$(c1ncccc1),\$(c1ccncc1)]	Polarised alkynes
Polarised alkyne pyrazine MA	[CH,\$(C(#C)[#6])#Cc1nccnc1	Polarised alkynes
Polarised alkyne pyrimidine MA	[CH,\$(C(#C)[#6])#C\$(c1ncccn1), \$(c1cncnc1),\$(c1ncncc1)]	Polarised alkynes
Polarised alkyne triazines MA	[CH,\$(C(#C)[#6])#C\$(c1ncncn1), \$(c1nncnc1),\$(c1nnccn1),\$(c1cnncn1)]	Polarised alkynes
Quinones MA	[\$(C1(=O)[#6]=,:[#6]C(=O)[CH]=C1), \$(C1(=O)C(=O)[#6]=,:[#6][CH]=C1)]	Quinones and quinone type chemicals
Quinone methides MA	[\$(C1=CC(=O)[#6]=,:[#6]C1(=[CH][CX4])), \$(C1(=[CH][CX4])C(=O)[#6]=,:[#6]C=C1)]	Quinones and quinone type chemicals
Quinone imines MA	[\$(C1(=O)[#6]=,:[#6]\$(C=[NH]),\$(C=N[CX4])) [CH]=C1),\$(C1(=O)[\$(C=[NH]),\$(C=N[CX4])) [#6]=,:[#6][CH]=C1),\$(C1(=O)[\$(C=[NH]), \$(C=N[CX4]))C=[CH][#6]=,:[#6]1)]	Quinones and quinone type chemicals
Quinone diimines MA	[\$([CH]1=C\$(C=[NH]),\$(C=N[CX4]))[#6]=,:[#6] \$(C=[NH]),\$(C=N[CX4]))1),\$([CH]1=C\$(C=[NH]), \$(C=N[CX4]))\$([C=[NH]),\$(C=N[CX4]))[#6]=,:[#6]1)]	Quinones and quinone type chemicals
Pyranones MA	[\$([cH]1cc(=[O,NH])cc[o,n]1), \$([cH]1cc(=[O,NH])[o,n]cc1), \$(c1cc(=[O,NH])[o,n][cH]c1)]	Quinones and quinone type chemicals
Acid imides MA	O=C1[NH,\$(N(C=O)(C=O)[#6])]C(=O) [CH,\$(C(C=O)(=C)[#6])]=[CH]1	Acid imides
Hydroquinones MA	[\$(c1([OH,\$(O[CH3]))c[cH]c([OH,\$(O[CH3]))]cc1), \$(c1([OH,\$(O[CH3]))c([OH,\$(O[CH3]))]cc[cH]c1)]	Pre quinones and quinone type chemicals
Alkyl phenols MA	[\$(c1([OH])ccc([CH3,\$([CH2][CX4]))cc1), \$(c1([OH])c([CH3,\$([CH2] [CX4]))cccc1)]	Pre quinones and quinone type chemicals
Aminophenols MA	[\$(c1([OH])c[cH]c([NH2])cc1), \$(c1([OH])c([NH2])cc[cH]c1)]	Pre quinones and quinone type chemicals
Benzenediamines MA	[\$(c1([NH2])c[cH]c([NH2])cc1), \$(c1([NH2])c([NH2])cc[cH]c1)]	Pre quinones and quinone type chemicals

Pre-polarised alkenes aldehydes (allyl alcohols) MA	<chem>[CH2,\$([CH](=C)[#6])]=[CH,\$(C(=C)([CH2][OH])[#6])[CH2][OH]</chem>	Pre polarised alkenes
Pre polarised alkynes aldehydes (propargyl alcohols) MA	<chem>[CH,\$(C(#C)[#6])#C[CH2][OH]</chem>	Pre polarised alkynes
Mono aldehydes SB	<chem>[CH2,\$([CH](=O)[CX4]),\$([CH](=O)c1a(*)caaa1),\$([CH](=O)c1a(*)aaa1)]=O</chem>	Direct acting Schiff base formers
Disubstituted polarised aldehydes SB	<chem>[\$(C(=C)([#6])[#6]);!\$(C(=C)([#6])C=[CH2,CH])]=C[CH]=O</chem>	Direct acting Schiff base formers
1,2-Dicarbonyl SB	<chem>[CH,\$(C(=O)(C=O)[CX4]),\$(C(=O)(C=O)c1a(*)aaaa1),\$(C(=O)(C=O)c1a(*)aaa1)](=O)[CH,\$(C(=O)(C=O)[#6]);!\$(C(=O)(C=O)C=[CH2,CH]);!\$(C(=O)(C=O)C#C)]=O</chem>	Direct acting Schiff base formers
1,3-Dicarbonyl SB	<chem>[CH,\$(C(=O)([CX4]C=O)[CX4]),\$(C(=O)([CX4]C=O)c1a(*)aaaa1),\$(C(=O)([CX4]C=O)c1a(*)aaa1)](=O)[CX4][CH,\$(C(=O)([CX4]C=O)[#6]);!\$(C(=O)([CX4]C=O)C=[CH2,CH]);!\$(C(=O)([CX4]C=O)C#C)]=O</chem>	Direct acting Schiff base formers
Ethylenediamines SB	<chem>[\$([NH2,\$([NH]([CH2])[CX4])[CH2][CH2][NH2]),\$([NH]1[CH2][CH2][NH][CH2][CH2]1)]</chem>	Pro Schiff base formers (glyoxal)
Ethanolamines SB	<chem>[\$([NH2,\$([NH]([CH2])[CX4])[CH2][CH2][OH]),\$([NH]1[CH2][CH2]O[CH2][CH2]1)]</chem>	Pro Schiff base formers (glyoxal)
Pro disubstituted polarised aldehydes (disubstituted allyl alcohols) SB	<chem>C([#6])([#6])=[CH,\$(C(=C)([CH2][OH])[#6])[CH2][OH]</chem>	Pro Schiff base formers (mono aldehydes)
Epoxides S _N 2	<chem>C1OC1</chem>	Epoxides and related
Aziridines S _N 2	<chem>C1[NX3]C1</chem>	Epoxides and related
Sulfaranes S _N 2	<chem>C1[SX2]C1</chem>	Epoxides and related
Epoxyp propane S _N 2	<chem>C1[CH2,CH]OC1</chem>	Epoxides and related
Epithiop propane S _N 2	<chem>C1[CH2,CH][SX2]C1</chem>	Epoxides and related
Lactones S _N 2	<chem>C1CC(=O)O1</chem>	Ring opening S _N 2 reaction
Nitrosoureas (nitrogen) S _N 2	<chem>[NX3]C(=O)[\$([NH](C=O)[NX2]=O),\$(N(C=O)([#6])[NX2]=O)]</chem>	S _N 2 reaction at a nitrogen atom
Nitrosoguanidine (nitrogen) S _N 2	<chem>[NX3]C(=[NH])[\$([NH](C=[NH])[NX2]=O),\$(N(C=[NH])([#6])[NX2]=O)]</chem>	S _N 2 reaction at a nitrogen atom
N-acetoxy N acetyl phenyl S _N 2	<chem>[a][NX3](O[\$([CH](O)=O),\$(C(=O)(O)[#6])])[\$([CH](N)=O),\$(C(=O)(N)[#6])]</chem>	S _N 2 reaction at a nitrogen atom
N-acyloxy N alkoxyamides S _N 2	<chem>[#6]O[NX3](O[\$([CH](O)=O),\$(C(=O)(O)[#6])])[\$([CH](N)=O),\$(C(=O)(N)[#6])]</chem>	S _N 2 reaction at a nitrogen atom

Isothiazol-3-ones (sulphur) S _N 2	[cH,\$(c(c)([sX2])([#6]))]1[cH,\$(c(c)(c=O)([#6]))c(=O)[nH,\$(n(c=O)([sX2])([#6]))][sX2]1	
Isothiazolin-3-ones (sulphur) S _N 2	[CH2,\$([CH](C)([SX2])([#6]))]1[CH2,\$([CH](C)(C=O)([#6]))C(=O)[NH,\$(N(C=O)([SX2])([#6]))][SX2]1	S _N 2 reaction at a sulphur atom
aromatic sulphonic acids S _N 2	[a][SX3](=O)[OH]	S _N 2 reaction at a sulphur atom
Thiocyanates S _N 2	[#6][SX2]C#N	S _N 2 reaction at a sulphur atom
Thiols S _N 2	[#6][SH]	S _N 2 reaction at a sulphur atom
Disulfides S _N 2	[#6][SX2][SX2][#6]	S _N 2 reaction at a sulphur atom
Thiosulfonates S _N 2	[#6][SX2][SX3](=O)[#6]	S _N 2 reaction at a sulphur atom
Sulfoxides of disulfides S _N 2	[#6][SX2][SX4](=O)(=O)[#6]	S _N 2 reaction at a sulphur atom
Sulfenyl halides S _N 2	[#6][SX2][F,Cl,Br,I]	S _N 2 reaction at a sulphur atom
N-chloro sulphonamides S _N 2	[#6][SX4](=O)(=O)[NH]Cl	S _N 2 reaction at a halo atom
N-haloimides S _N 2	[CH,\$(C(=O)(N)[#6]))(=O)[NX3]([F,Cl,Br,I])[CH,\$(C(=O)(N)[#6]))=O	S _N 2 reaction at a halo atom
Polarised alkenes with a halogen leaving group S _N 2	C=[CH][F,Cl,Br,I]	S _N 2 reaction at a sp ² carbon
Polarised alkenes with a sulfonate leaving group S _N 2	C=[CH]OS(=O)(=O)[#6]	S _N 2 reaction at a sp ² carbon
Polarised alkenes with a sulfate leaving group S _N 2	C=[CH]OS(=O)(=O)O[#6]	S _N 2 reaction at a sp ² carbon
Polarised alkenes with a phosphonate leaving group S _N 2	C=[CH]OP(=O)([O,SX])(O[#6])[#6]	S _N 2 reaction at a sp ² carbon
Polarised alkenes with a phosphate leaving group S _N 2	C=[CH]OP(=O)([O,SX])(O[#6])O[#6]	S _N 2 reaction at a sp ² carbon
Mustards S _N 2	[NX3,SX2][CH2,\$([CH]([NX3,SX2])([CX4])([CX4]))[CH2,\$([CH]([CX4])([Cl,Br,I])([CX4]))[Cl,Br,I]	Episulfonium ion formation
1,2-Dihaloalkanes S _N 2	[Cl,Br,I][CH2,\$([CH]([Cl,Br,I])([CX4])([CX4]))[CH2,\$([CH]([Cl,Br,I])([CX4])([CX4]))[Cl,Br,I]	Episulfonium ion formation
Aliphatic halide S _N 2	[CH3,\$([CH2]([F,Cl,Br,I])([#6])), \$([CH]([F,Br,Cl,I])([#6])([#6])); !\$([CH2,CH]C=O); !\$([CH2,CH][CX4][NX3,SX2]); !\$([CH2,CH][CX4][F,Cl,Br,I])[F,Cl,Br,I]	S _N 2 at an sp ³ carbon atom
Sulfates S _N 2	[#6]OS(=O)(=O)O[CH3,\$([CH2](O)[#6]), \$([CH](O)([#6])([#6]))]	S _N 2 at an sp ³ carbon atom
Sulfonates S _N 2	[#6]S(=O)(=O)O[CH3,\$([CH2](O)[#6]), \$([CH](O)([#6])([#6]))]	S _N 2 at an sp ³ carbon atom
Allyl acetates and related S _N 2	[#6]C(=[O,SX1])[O,SX2][CH2,\$([CH]([O,SX2])([#6]))][c,CX3,CX2]	S _N 2 at an sp ³ carbon atom

Nitrosoureas (carbon) S _N 2	[NH2,\$([NH](C=[O,NH])[#6]), \$(N(C=[O,NH])([#6])([#6]))C(=[O,NH]) [NX3]([NX2]=O)[CH3,\$([CH2](N)[#6]), \$([CH](N)([#6])([#6]))]	S _N 2 at an sp ³ carbon atom
α-Halocarbonyls S _N 2	[\$([CH]=[O,SX1]),\$(C([#6])=[O,SX1])] [CH2,\$([CH](C=[O,SX1])([#6]))][F,Cl,Br,I]	S _N 2 at an sp ³ carbon atom
Phosphates (inc thiophosphates) S _N 2	[#6][O,SX2]P(=[O,SX1])([O,SX2][#6]) [O,SX2][CH3,\$([CH2](O)[#6]),\$([CH](O)([#6])([#6]))]	S _N 2 at an sp ³ carbon atom
Phosphonates (inc thiophosphonates) S _N 2	[#6][O,SX2]P(=[O,SX1])([#6][O,SX2] [CH3,\$([CH2](O)[#6]),\$([CH](O)([#6])([#6]))]	S _N 2 at an sp ³ carbon atom
α-Haloethers S _N 2	[CX4]O[CH2,\$([CH](O)([F,Cl,Br,I])([#6]))][F,Cl,Br,I]	S _N 2 at an sp ³ carbon atom
β-Haloethers S _N 2	[CX4]O[CH2,\$([CH](O)([CX4])([CX4]), \$(C(O)([CX4])([CX4])([CX4]))[CH2, \$([CH](O)([F,Cl,Br,I])([#6]))][F,Cl,Br,I]	S _N 2 at an sp ³ carbon atom
Alkyl diazo S _N 2	[CH3,\$([CH2](N=N)[#6]),\$([CH](N=N) ([#6])([#6]))[NX2]=[NH,\$(N=N)[#6]]]	S _N 2 at an sp ³ carbon atom
α-Haloalkenes S _N 2	[CH2,\$([CH](=C)[#6]),\$(C(=C)([#6]) [#6])]=[CH,\$(C(=C)[#6])][\$([CH2](C=C) [F,Cl,Br,I,\$(C#N),\$(OS(=O)(=O)O[#6]), \$(OS(=O)(=O)[#6]))],\$([CH](C=C)([#6]) [F,Cl,Br,I,\$(C#N),\$(OS(=O)(=O)O[#6]), \$(OS(=O)(=O)[#6]))]	S _N 2 at an sp ³ carbon atom
α Haloalkynes S _N 2	[CH,\$(C#C)[#6])#C[\$([CH2](C#C) [F,Cl,Br,I,\$(C#N),\$(OS(=O)(=O)O[#6]), \$(OS(=O)(=O)[#6]))],\$([CH](C#C)([#6]) [F,Cl,Br,I,\$(C#N),\$(OS(=O)(=O)O[#6]),\$(OS(=O)(=O)[#6]))]	S _N 2 at an sp ³ carbon atom
α-Halobenzyls S _N 2	[c][\$([CH2]([c])[F,Cl,Br,I,\$(C#N), \$(OS(=O)(=O)O[#6]),\$(OS(=O)(=O)[#6]))], \$([CH]([a])([#6])[F,Cl,Br,I,\$(C#N), \$(OS(=O)(=O)O[#6]),\$(OS(=O)(=O)[#6]))]	S _N 2 at an sp ³ carbon atom
Activated benzenes S _N Ar	[\$(c1([F,Cl,Br,\$(OC(F)(F)F), \$(C#N)])c([N+](=O)[O-])cc([N+](=O)[O-])cc1,\$(c1(F)c([N+] (=O)[O-])cc([\$([CH]=O),\$(C(=O)[CH3]),\$(C(F) (F)F))cc1,\$(c1(Cl)c([N+] (=O)[O-])cc([CH]=O)cc1,\$(c1(F)c([\$(C#N),\$([CH]=O))cc([N+] (=O)[O-])cc1,\$(c1([\$(OC(F)(F)F),Br,\$(C#N)])c([N+] (=O)[O-])cc([CH]=O)cc1,\$(c1([F,Cl,Br,\$(OC(F) (F)F),\$(C#N)])cc([N+](=O)[O-])c([N+] (=O)[O-])cc1,\$(c1([F,Cl,Br,\$(OC(F)(F)F),\$(C#N)])c([N+] (=O)[O-])cccc1([N+](=O)[O-]))]	Activated benzenes
Activated poly fluorobenzenes S _N Ar	[\$(c1([\$([N+](=O)[O-]),\$([CH]=O),\$(C#N),\$(C(=O) [CH3]))c(F)c(F)c(F)c(F)c1(F)),\$(c1([\$([N+] (=O)[O-]),\$([CH]=O),\$(C#N)])c(F)c(F)c(F)c(F)c1,\$(c1([\$([N+] (=O)[O-]),\$([CH]=O),\$(C#N)])c(F)c(F)c(F)cc1F,\$(c1([\$([N+] (=O)[O-]),\$([CH]=O),\$(C#N)])c(F)c(F)cc(F)c1F,\$(c1([N+] (=O)[O-])c(F)c(F)c(F)cc1,\$(c1([N+] (=O)[O-])c(F)c(F)cc(F)c1,\$(c1([N+] (=O)[O-])c(F)c(F)ccc1F,\$(c1([N+](=O)[O-])c(F)cc(F)cc1F)]	Activated poly fluorobenzenes
Activated pyridines S _N Ar	[\$(c1([F,Cl,Br,\$(C#N),\$(OC(F) (F)F))ncc([\$(N(=O)=O),\$(C#N),\$([CH]=O),\$(C(=O) [CH3]))cc1,\$(c1([F,Cl,Br,\$(C#N),\$(OC(F) (F)F))ncccc1[\$(N(=O)=O),\$(C#N),\$([CH]=O),\$(C(=O) [CH3]))],\$(c1([F,Cl,Br,\$(C#N),\$(OC(F) (F)F))ccncc1[\$(N(=O)=O),\$(C#N), \$([CH]=O),\$(C(=O)[CH3]))]	Activated pyridines

Procedure Description

Biological Endpoint and Measurement:

This method is an *in silico* model, which provides information on the following MIE, biologically relevant, for several toxicological endpoints:

Covalent binding to proteins

Endpoint Value:

Qualitative: Structural alert

Experimental System:

In silico profiler for covalent protein binding

3. Method Performance

Robustness of the Method

General performance measures:

The applicability domain of the structural alerts obtained with the profilers was investigated for different groups of chemicals using the glutathione (GSH) reactivity assay and the *Tetrahymena pyriformis* growth inhibition assay:

- Potential pre-electrophile polyphenolics were evaluated by Bajot et al. (2010)
- Epoxides, lactones, nitrosos, nitros, aldehydes, ketones and ring-strained hydrocarbons were evaluated by Nelms et al. (2013)
- Cyclic compounds as potential Michael acceptor were evaluated by Rodriguez-Sanchez et al. (2013)
- Sulfur-containing organic chemicals were evaluated by Richarz et al. (2014)

Each of the studies showed significant agreement between the profilers and experimental data, however several discrepancies were observed, which were explained once additional mechanistic information was taken into account. Further refinement of the profiler was recommended for each case, underpinning the importance of using extensive experimental data to further improve the accuracy of the alerts.

4. Discussion

The formation of a covalent adduct between a reactive pair of exogenous chemical (electrophile) and biological protein (nucleophile) is one of the fundamental steps for several toxicological endpoints such as hepatotoxicity (liver fibrosis), skin sensitisation, respiratory sensitisation or aquatic toxicity. The mechanistic importance of this chemical reaction makes the mechanistic alert approach the natural choice for developing the profiler for covalent protein binding. Moreover the mechanistic domain overlap between corresponding structural alerts in the literature compilations has been assessed. This analysis ensured that for a given structural alert the maximum mechanistic information (and thus domain) was extracted.

Unlike covalent DNA binding, the covalent protein binding is not distinctly linked to one specific area of toxicology. Therefore, no established experimental toxicological dataset exists that could be readily used in the verification of structural alerts and for applicability domain definition. The interpretation of the alerts will largely depend on the endpoint for which the chemical is being considered and corroborating information available for that endpoint. The alerts, however, can be verified using *in chemico* data, e.g.

reactivity with GSH.

The use of such mechanistic chemistry based alerts enables transparent category formation where all category members act via the same mechanism of action (OECD 2007; ECHA 2008). Moreover, such alerts, accompanied by the detailed documentation of the chemistry relevant to covalent protein binding, make the chemical category methodology acceptable for safety assessment without animal testing.

5. Contact Details

Dr Steve Enoch

School of Pharmacy and Biomolecular Sciences
 Liverpool John Moores University
 Corresponding author, Method developer / user
 Byrom Street
 Liverpool L3 3AF
 email: s.j.enoch@ljmu.ac.uk
 telephone: +44(0)151 231 2164

Ver n°	Date	Name	Type of change
1	12 November 2015	Dr Steve Enoch	Data sheet created

6. Abbreviations & Definitions

AC Acylation

AOP Adverse Outcome Pathway

COSMOS EU Integrated Project: COSMOS developed an integrated suite of computational tools to predict the effects of long-term exposure to chemicals in humans based on *in silico* calculation and estimation techniques.

Website: <http://www.cosmostox.eu>.

ITS Integrated Testing Strategy

IATA Integrated Approach to Testing and Assessment

KNIME KoNstanz Information MinEr

MA Michael Addition

MIE Molecular Initiating Event

SB Schiff Base formation

SEURAT Safety Evaluation Ultimately Replacing Animal Testing

SMARTS SMiles ARbitrary Target Specification

SMILES Simplified Molecular-Input Line-Entry System

S_N2 Bimolecular Aliphatic Nucleophilic Substitution

S_NAr Aromatic Nucleophilic Substitution

7. Bibliography

Method development:

- Bajot F., Cronin M.T., Roberts D.W., Schultz T.W. (2011)
 Reactivity and aquatic toxicity of aromatic compounds transformable to quinone-type Michael

acceptors.

SAR and QSAR in Environmental Research 22(1-2), 51-65

- Enoch S.J., Cronin M.T.D., and Ellison C.M. (2011)
The use of a chemistry-based profiler for covalent DNA binding in the development of chemical categories for read-across for genotoxicity.
Alternatives to Laboratory Animals (ATLA) 39, 131-145
- Enoch S.J., Ellison C.M., Schultz T.W., Cronin M.T.D. (2011)
A review of the electrophilic reaction chemistry involved in covalent protein binding relevant to toxicity.
Critical Reviews in Toxicology 41, 783-802
- Nelms M.D., Cronin M.T.D., Schultz T.W., and Enoch S.J. (2013)
Experimental verification, and domain definition, of structural alerts for protein binding: epoxides, lactones, nitroso, nitros, aldehydes and ketones.
SAR and QSAR in Environmental Research 24, 695-709
- Richarz A.N., Schultz T.W., Cronin M.T.D., and Enoch S.J. (2014)
Experimental verification of structural alerts for the protein binding of sulfur-containing compounds.
SAR and QSAR in Environmental Research 25, 325-341
- Rodriguez-Sanchez N., Schultz T.W., Cronin M.T.D., and Enoch S.J. (2013)
Experimental verification of structural alerts for the protein binding of cyclic compounds acting as Michael acceptors.
SAR and QSAR in Environmental Research 24, 963-977

General references:

- Chipinda I., Hettick J.M., Siegel P.D., Chipinda I., Hettick J.M., and Siegel P.D. (2011)
Haptenation: Chemical Reactivity and Protein Binding, Haptenation: Chemical Reactivity and Protein Binding.
Journal of Allergy and Clinical Immunology
- Cochrane S.A., Arts J.H.E., Ehnes C., Hindle S., Hollnagel H.M., Poole A., Suto H., and Kimber I (2015)
Thresholds in chemical respiratory sensitisation.
Toxicology 333, 179-194
- ECHA (2008)
Guidance on information requirements and chemical safety assessment Chapter R.6: QSARs and grouping of chemicals. Available from: http://guidance.echa.europa.eu/docs/guidance_document/
- EU (2009)
Regulation (EC) No 1223/2009 of the European Parliament and of the Council of 30 November 2009 on cosmetic products. [Link to document](#) (last access 13.09.2016)
Official Journal of the European Communities
- EU (2006)
Regulation (EC) No 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH). [Link to document](#) (last access 16.09.2013)
Official Journal of the European Communities L396
- Landesmann B. (2016)
AOP 38: Protein Alkylation leading to Liver Fibrosis. Access: <https://aopwiki.org/wiki/index.php/Aop:38>
- Landesmann B. (2016)
AOP 40: Covalent Protein binding leading to Skin Sensitisation. Access: <https://aopwiki.org/wiki/index.php/Aop:40>
- Landesmann B. (2012)
Description of Prototype Modes-of-Action Related to Repeated Dose Toxicity,
[JRC publication repository](#) EUR 25631 EN
- Lopachin R.M., Decaprio A.P. (2005)
Protein adduct formation as a molecular mechanism in neurotoxicity.
Toxicological Sciences 86, 214-225
- OECD (2007)
Guidance on Grouping of Chemicals
[OECD Environmental Health and Safety Publications](#)
- Zhou D.S., Chan E., Duan W., Huang M., and Chen Y.-Z. (2005)

Drug Bioactivation Covalent Binding to Target Proteins and Toxicity Relevance.
Drug Metabolism Reviews 37, 41–213

Last update: 12 November 2015

DB-ALM Method summary n° 182 : *In silico* skin permeability estimation

Systemic Toxicity

The purpose of this *in silico* workflow is to estimate the skin permeability coefficient (kp) for an organic compound, based on calculated physico-chemical descriptors.

1. General Information

Corresponding author(s):

Prof. Mark Cronin, Liverpool John Moores University

Date of data sheet creation: 28 October 2015

Last version number: 1

Last update: 28 October 2015

Proprietary and/or Confidentiality Issues:

None. The model is publicly available.

Status

Development

Development completed. Work carried out at Liverpool John Moores University with support from the FP7 COSMOS Project (Integrated *In Silico* Models for the Prediction of Human Repeated Dose Toxicity of Cosmetics to Optimise Safety), within the SEURAT-1 Research Initiative (Safety Evaluation Ultimately Replacing Animal Testing 1) cluster (2011-2015). More information on the COSMOS project is available on its dedicated website:

<http://www.cosmostox.eu>



The *in silico* model has been coded into a KNIME (<http://www.knime.org/>) automated workflow as part of the COSMOS project. The web version of the KNIME workflow can be executed via a web browser. It is available through the COSMOS KNIME WebPortal:

<http://knimewebportal.cosmostox.eu>

More information is available via COSMOS Space:

<http://cosmosspace.cosmostox.eu>

A tutorial is available at:

<https://www.youtube.com/watch?v=DtNdwTKuY74>

2. Method Definition

Purpose of the Method

TYPE OF TESTING : Screening

Context of use

Dermal absorption is of great importance for safety evaluation of chemicals taken up via dermal route of exposure regarding the availability of the substance in the organism. Experimental testing of the multitude of chemical substances for properties such as skin absorption is laborious and expensive, even with the use of *in vitro* methods e.g. such as described in the **DB-ALM Topic Summary "Percutaneous Absorption"** (DB-ALM, 2010). Therefore, reliable models for a theoretical prediction of dermal absorption have been sought (Cronin, 2006; Russel and Guy, 2009).

An open-access skin permeability estimation automated model is presented to support risk assessment and safety evaluation regarding dermal absorption potential. The purpose of this model is to obtain skin permeability coefficient estimations with an easy and fast but also robust method.

The model was built to support safety evaluation of cosmetics-related substances. The robustness of the model by incorporating statistical data quality is beneficial for the chemical diversity of cosmetic ingredients. Furthermore the model might support evaluations in the context of REACH (EU, 2006), medical devices etc. Principally the model may support other fields as well, e.g. dermal drug delivery, for example within a consensus approach with known data.

Scientific Principle of the Method

Substances with high octanol/water partition coefficients (a common measure of lipophilicity) and low molecular volume are associated with greater skin permeability coefficient (k_p) and hence greater dermal absorption.

Many models, most notably the Quantitative Structure-Property Relationship (QSPR) of Potts and Guy (1992), have been created to predict skin permeability coefficients (k_p) or flux (J) of dermally applied substances.

Similar computational models, i.e. correlating physio-chemical properties with membrane permeability, are used widely in different scientific fields. These models are developed and used in the investigation and assessment of pharmaco- and toxicokinetics. They are particularly relevant for oral and dermal drug development. However, they are also valuable contributions to the safety evaluation of cosmetic ingredients.

Model description

The Potts and Guy QSPR (1992) to predict skin permeability was modified by incorporating a larger dataset and a statistical tool to assess data quality (confidence score).

The compilation of skin permeability coefficient values from the literature (see below) resulted in 343 different k_p values being compiled for 226 compounds. 55 of these compounds have more than a single k_p value, hence the arithmetic means and the confidence scores (CS) were calculated. Due to the high variability in most skin permeability datasets, a statistical data quality tool was used (Steinmetz et al., 2014). The assigned statistical data quality was used to support weighted linear regression. Hence the QSPR equation is based on CS-weighted multivariate linear regression.

Physico-chemical properties, i.e. molecular volume (MV) and lipophilicity (XLogP) were calculated from molecular structures, expressed as SMILES (Simplified Molecular-Input Line-Entry System) strings, with CDK in KNIME 2.9.1. R Studio 0.98.501 was used to model k_p with MV and XLogP as descriptors and CS as weights. The results of the multivariate linear regression is expressed in Equation 1 below.

The model was implemented as KNIME workflow web version, as an easy-to-use user-interface allowing to execute the model via a web browser. It is available through the COSMOS KNIME WebPortal:

<http://knimewebportal.cosmostox.eu>

Dataset

Data was collected from the following literature:

- ten Berge (2014)
- Chen et al. (2013)
- Moss and Cronin (2002)
- Chauhan and Shakya (2010).

Model equation

The equation (Eq.1) used within the skin permeability estimation KNIME workflow is as follows:

$$\text{Log } k_p = -2.33 + 0.513 \text{ XLogP} - 0.00611 \text{ MV (Eq. 1)}$$

$$n = 226, S_{\text{XLogP}} = 0.035, t_{\text{XLogP}} = 14.8, S_{\text{MV}} = 0.000685, t_{\text{MV}} = -8.91, F = 113.03, R^2_{\text{adj}} = 0.499$$

Procedure Description

Biological Endpoint and Measurement:

This method is an *in silico* model, which provides information on the following, biologically relevant, endpoint:

Dermal absorption

Endpoint Value:

Human *in vitro* skin permeability coefficient (k_p)

Experimental System:

In silico workflow to estimate the human skin permeability coefficient

3. Method Performance

Robustness of the Method

General performance measures:

The QSPR model was validated with 10-fold cross validation employing confidence score (CS)-adjusted root-mean-square error (RMSE) (Eq.2), which led to a value of 0.79 ± 0.25 . Further details on method validation can be found in Steinmetz et al. (2015).

$$RMSE_{CS} = \sqrt{\frac{\sum_i CS_i (f_i - y_i)^2}{\sum_i CS_i}} \quad (\text{Eq.2})$$

Test Compounds

Applicability:

The model is only applicable for simple organic molecules, i.e. no mixtures, salts or polymers are supported.

4. Discussion

The robust estimation of human *in vitro* k_p values may support regulatory decision making. It is however not recommended to use this value as single rationale/argument in safety evaluations based on dermal absorption.

Ethical issues and considerations for 3R's impact

There is an indirect positive impact on 3Rs due to the development of a new methodology (implementing statistical data quality into QSAR/QSPR modelling). Furthermore this model supports integrated testing strategies and non-testing approaches.

Strengths and limitations of the method

It must be emphasised that the modelling approach used here is rather simple and hence can be repeated with appropriate data and descriptors. Performance has been compromised by using freeware tools from CDK within KNIME. A very similar model with slightly better performance using different descriptors has been published in Steinmetz et al. (2015).

Since the predictivity of the model is only moderate, the model should be used for support or estimation only. However, based on its transparency and implementation of data quality, the model can be regarded as more robust than many overpredicted models (for example due to defined chemical categories). EPI Suite's Dermwin, just to mention one of many models, predicts *kp* values as well. For well-studied chemical classes, i.e. where many *kp* data is available, models such as Dermwin are likely to perform better. However, for new chemical classes, the described model is assumed to perform better, based on its robustness over different chemical classes with a pure focus on physico-chemical properties.

Potential for future development

The implementation of new data (if available) is recommended. Building models with other descriptors and/or only specific chemical compounds is possible. If such models are desired, please refer to current QSAR guidelines.

5. Contact Details

Dr Fabian Steinmetz

School of Pharmacy and Biomolecular Sciences
Liverpool John Moores University
Method developer / user
Byrom Street
Liverpool L3 3AF

Prof. Mark Cronin

Professor of Predictive Toxicology
School of Pharmacy and Biomolecular Sciences
Liverpool John Moores University
Corresponding author
Byrom Street
Liverpool L3 3AF
email: m.t.cronin@ljmu.ac.uk
telephone: + 44 151 231 2402
fax: + 44 151 231 2170
last contact update: 11 May 2016

Ver n°	Date	Name	Type of change
1	28 October 2015	Prof. Mark Cronin	Datasheet created

6. Abbreviations & Definitions

CDK: Chemical Development Kit

COSMOS: EU Integrated Project: COSMOS developed an integrated suite of computational tools to predict the effects of long-term exposure to chemicals in humans based on *in silico* calculation and estimation techniques.

Website: <http://www.cosmostox.eu>

CS: Confidence Score

MV: Molecular Volume (in Å³)

QSAR: Quantitative Structure-Activity Relationship

QSPR: Quantitative Structure-Property/-Permeability Relationship

RMSE: Root-Mean-Square Error

SEURAT-1: Safety Evaluation Ultimately Replacing Animal Testing

SMILES: Simplified Molecule Input Line Entry System

XLogP: CDK's calculated octanol/water partition coefficient

7. Bibliography

Method development:

- Chauhan P, Shakya M. (2010)
Role of physicochemical properties in the estimation of skin permeability: in vitro data assessment by Partial Least-Squares Regression.
SAR and QSAR in Environmental Research 21, 481-494
- Chen L et al. (2013)
Recent advances in predicting skin permeability of hydrophilic solutes.
Advanced Drug Delivery Reviews 65, 295-305
- Moss, G. P., Dearden, J. C., Patel, H., Cronin, M. T. (2002)
Quantitative structure-permeability relationships (QSPRs) for percutaneous absorption.
Toxicology In Vitro 16, 299-317
- Potts, R. O., Guy, R. H. (1992)
Predicting skin permeability.
Pharmaceutical Research 9, 663-669
- Steinmetz, F.P., Enoch, S.J., Madden, J.C., Nelms, M.D., Rodriguez-Sanchez, N., Rowe, P.H., Wen, Y., Cronin, M.T.D. (2014)
Methods for assigning confidence to toxicity data with multiple values – identifying experimental outliers.
Science of the Total Environment 482, 358-365
- Steinmetz, F.P., Madden, J.C., Cronin, M.T.D., (2015)
Data Quality in the Human and Environmental Health Sciences: Using Statistical Confidence Scoring to Improve QSAR/QSPR Modeling
Journal of Chemical Information and Modeling 55, 1739-1746
- ten Berge, W (2014)
<http://home.wxs.nl/~wtberge/skinperm2013a.zip> (accessed 01.03.2014)

General references:

- Cronin, M. T. D. (2006)
Chapter title: The prediction of skin permeability using quantitative structure-activity relationship methods. Book title: Dermal Absorption Models in Toxicology and Pharmacology, by J. E. Riviere (ed.). CRC Press-Taylor & Francis Group, Boca Raton, FL, S. 113-134.
- DB-ALM (2010)
Topic Summary "Percutaneous Absorption"
[*EURL ECVAM Database service on Alternative Methods to animal experimentation*](#)
- EU (2006)
Regulation (EC) No 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), establishing a European Chemicals Agency, amending Directive 1999/45/EC and repealing Council Regulation (EEC) No 793/93 and Commission Regulation (EC) No 1488/94 as well as Council Directive 76/769/EEC and Commission Directives 91/155/EEC, 93/67/EEC, 93/105/EC and 2000/21/EC. [Link to document](#) (last access 16.09.2013)
Official Journal of the European Communities L396
- Mitragotri S et al (2011)
Mathematical models of skin permeability: an overview.
International Journal of Pharmaceutics 418, 115-129
- Russell, L. M., Guy, R. H. (2009)
Measurement and prediction of the rate and extent of drug delivery into and through the skin.
Expert Opinion on Drug Delivery 6, 355-369
- US EPA (2015)
EPI Suite software; <http://www2.epa.gov/tsca-screening-tools> (accessed 10.10.2015)

Last update: 28 October 2015

Europe Direct is a service to help you find answers to your questions about the European Union

Free phone number (*): 00 800 6 7 8 9 10 11

(*) Certain mobile telephone operators do not allow access to 00 800 numbers or these calls may be billed.

A great deal of additional information on the European Union is available on the Internet.

It can be accessed through the Europa server <http://europa.eu>

How to obtain EU publications

Our publications are available from EU Bookshop (<http://bookshop.europa.eu>), where you can place an order with the sales agent of your choice.

The Publications Office has a worldwide network of sales agents.
You can obtain their contact details by sending a fax to (352) 29 29-42758.

JRC Mission

As the science and knowledge service of the European Commission, the Joint Research Centre's mission is to support EU policies with independent evidence throughout the whole policy cycle.



EU Science Hub

ec.europa.eu/jrc



@EU_ScienceHub



EU Science Hub - Joint Research Centre



Joint Research Centre



EU Science Hub